Neural Substrates of Choosing Actions and Motivational Drive, a Role for the Striatum

Dissertation advisor: Dr. Naoshige Uchida Alice Yiqing Wang

Neural substrates of choosing actions and motivational drive, a role for the striatum

Abstract

Optimal decision making requires one to determine the best action among available alternatives as well as the most appropriate level of engagement for performance. While current research and models of decision making have largely focused on the former problem, or action selection, less is known about the latter problem of the selection of motivational drive. Thus, I designed a self-paced decision-making paradigm that aimed to dissociate both facets of selection in rats. First, I showed that the expected net value of potential options influenced rats' general motivation to perform: rats globally exhibited shorter latency to initiate trials in states of high net return than in states of low net return. In contrast, the relative value of options biased choice direction. To study the neural substrates underlying either process, I examined the role of the striatum, which is closely connected with cortex and dopamine neurons, acting as a major hub for reward-related information. In chapter 1, I show that selective lesions of the dorsomedial (DMS) but not ventral striatum (VS) impaired net value-dependent motivational drive but largely spared choice biases. Specifically, DMS lesions rendered animals' latency to initiate trials dependent on the absolute value of immediately preceding trial outcomes rather than on the net value of options. Accordingly, tetrode recordings in Chapter 2 showed that the DMS rather than VS predominantly encodes net value. In fact, net value representation in the DMS was stronger than either absolute or

relative value representations during early trial epochs. Thus, the DMS flexibly encodes net expected return, which can guide the selection of motivational drive.

Table of Contents

Table of Figures

Acknowledgements

I am extremely fortunate to have been a part of the excellent neuroscience community at Harvard, which spans the Department of Neurobiology, the Center for Brain Science, and the Department of Molecular and Cellular Biology. I would like to acknowledge members of my DAC committee, Drs Marge Livingstone, John Maunsell, and Rachel Wilson for their support and guidance throughout my PhD. Further, I would like to thank the members of my thesis committee, Drs Richard Born, Okihide Hikosaka, and Daeyeol Lee, and the chair, Marge Livingstone.

I thank the all the past and present members of the Uchida lab, particularly Keiji Miura, with whom I collaborated in Chapter 2, Methods. Finally, I am indebted to my advisor, Dr. Naoshige Uchida, who provided extraordinary mentorship, knowledge, enthusiasm and inspiration throughout my time at Harvard.

Lastly, I thank my parents, Igor, my aunt, and my grandparents for their everlasting encouragement. This thesis is dedicated to my grandfather, Professor Kaiqing Yang. This work is supported by the National Science Foundation's Graduate Research Fellowship Program.

INTRODUCTION

A fundamental function of the nervous system is to select behaviors that maximize the chances of survival. This requires continual monitoring of the internal state and goals of the animal as well as the opportunities or dangers present in the environment. As both internal and external conditions fluctuate dynamically and often unpredictably, the selection of behaviors must be performed quickly and flexibly to adapt to the demands of the environment and the animal's individual needs. How to decide the optimal behavior to perform for a given situation is therefore a key question not limited to neuroscience, psychology, economics, and machine learning.

Optimal decision making can be broken into multiple sub-choices such as choosing what action to do, how to do it, or whether to perform at all. These sub-choices are often performed simultaneously. To date, advances in computational, behavioral, and electrophysiological neurosciences have contributed significantly to our understanding of multiple aspects of decision making. Below, I review major conceptual and theoretical contributions to two aspects of decision making: the selection of motivational drive and the selection of specific actions.

Models of Selection: choosing motivational drive and actions

Motivation

Generalized drive

Perhaps the most commonplace observation of animals dynamically adapting their behavioral strategies according to their needs is that when animals are thirsty they are motivated to drink, and when they are hungry they are driven to eat. Among the earliest influential theories used to explain motivated behavior is Clark Hull's theory of drive reduction (Hull 1943). In this theory, the behavioral output of an animal or 'reaction potential' is a function of the animal's general drive and its learned habits:

$$Reaction\ potential = {}_{S}H_{R} \times D$$

Hull is most famous for introducing the generalized drive, or *D*, which is 'singular, not plural, general, not specific, and motivational, not directing,' thus energizing behavior in a non-specific way (Bolles 1975). Various needs such as food or water deprivation constitute different ways of activating drive, but drive itself is singular. Moreover, he postulated that the internal buildup of drive is aversive, and therefore its reduction is reinforcing, acting as the chief mechanism for reward. While drive is used to motivate behavior, it cannot be used to steer behavior. In order to 'direct' actions toward specific outcomes, Hull included 'specific drives', which are guided by stimulus-response habits that are learned through reinforcement, or ${}_{s}H_{R}$. In his equation, the reaction potential of an animal critically depends

on the presence of both the general and specific drive: if either D or sH_R is zero, the animal does not perform.

While drive energizes behavior in a non-specific way, specific drive directs behavior towards actions that have been learned to yield specific outcomes. This model can explain why hungry rats are directed towards food, but simultaneously, they are generally more 'jumpy', such as grooming or running on exercise wheels faster or more vigorously. Supporters of drive theory included Kenneth Spence and Donald Hebb, who famously drew an analogy that behavior requires an engine as well as a steering wheel, one for energizing and the other for directing behavior (Hebb 1955).

The notion of the generalized drive D, in particular, has been extremely controversial and has largely fallen out of use. Prominent figures in psychology such as Bolles argued that specific drive alone may be sufficient to explain motivated behaviors, as 'by the time we have discovered what the associative determinants are for any particular behavior, there is very little left for [drive] to explain' (Bolles 1975). Therefore, the need to introduce a general drive variable may simply reflect the experimenter's ignorance of the stimulus that had activated the observed response. Further, over several decades, many experimental observations have contradicted aspects of Hull's drive reduction theory. For example, feeding nutrients intravenously or intragastrically should reduce drive buildup, yet it does not prevent eating (Nicholaidis and Rowland 1974). Furthermore, it could not explain why animals and people abuse of drugs of addiction, or explain the powerfully rewarding effects of electrical self-stimulation of certain brain areas (Olds and Milner 1954). Finally, although

the concept of one singular drive may be appealing, 'the worst failure of the drive concept is that it provides little help in explaining behavior' (Bolles, 1975).

Schedules of reinforcement and the choice of response rate

Opposite to Hull, Skinner approached the study of motivation entirely based on empirical observations. One of Skinner's key contributions was that response rates and motivation levels of animals can be controlled by schedules of reinforcement imposed by the experimenter. Specifically, he observed that different types of reinforcement schedules controlled by the experimenter could profoundly influence animal's motivation or vigor. In his free-operant tasks, animals were free to choose their rate of responding, and behaviors were measured by counting the number or strength of lever presses or key taps performed by rats or pigeons (Ferster and Skinner 1957). For example, experimenters trained animals to perform tasks under ratio or interval schedules. In ratio schedules, reinforcement depends on the *number* of responses performed while in interval schedules, reinforcement becomes available after a certain amount of *time* has elapsed. Further, interval or ratio schedules could be fixed or variable. In fixed ratio schedules, the ratio of responses to reward remain fixed (e.g. a 'FR5' schedule means reward is given for every 5 responses). In fixed interval schedules, reinforcement is given at fixed time intervals (e.g., 'FI2: reward is given every 2 seconds). Animals showed faster rates of responding in ratios schedules, where reinforcement depended on the number of responses than in interval schedules, where reinforcement depended only on the passage of time. Strikingly, this phenomenon held true even if the frequency of reinforcement was the same in both fixed and interval schedule tasks

(yoked interval schedules). Moreover, response patterns were different. In fixed ratio tasks, responses were quite regular whereas in fixed interval schedules, animals reserved responses until the deadline of the interval, implying that animals may compute a trade-off between reward with the energetic costs of performance.

In contrast, variable ratio and variable interval schedules appeared to significantly increase response rates compared with their fixed ratio and fixed interval counterparts. In these schedules, the number of responses to get reward (in the case of variable ratio) or the number of time elapsed before reward (in the case of variable interval) varies from trial to trial, around some mean. For example, a schedule of 'VR5', reward is given *on average* every 5 responses and or with VI5, reward is given *on average* every 5 seconds. Even if the schedules were matched such that animals experienced the same rate of reward, variable schedules elicited higher rates of responding than their fixed schedule counterparts. *Variable ratio* schedules, in particular, produced the highest rate of responding, but both variable ratio and interval schedules were resistant to extinction. Taken together, these seminal studies on the effects of schedules of reinforcement on response rates are among the first evidence that animals select their response rates according to tradeoffs among many factors such as reward, energetic costs, uncertainty, etc.

Herrnstein, a student of Skinner's, formulated a quantitative relationship between the rate of reinforcement and response rates in his 'matching law' (Herrnstein 1970),

$$P1/(P1+P2) = R1/(R1+R2)$$

where P1 and P2 is the rate of pecking a left and right key, respectively, and R1 and R2 are the respective rates of reinforcement for pecking the keys. Given two concurrently available but different reinforcement schedule options, he found that animals will match the ratio of responding on the two levers to the ratio of their experienced reward rates. For example, if a left lever has a schedule of VR5 while a right lever has a schedule of VR10, the animal will press the right lever twice as fast as the left. Thus, animals can temporally regulate their behavior as a means to maximize long-term rewards in a predictable and quantifiable way. In addition to explaining response rates, the matching law offered one of the earliest models for predicting action selection (i.e. the choice of *which* lever to press).

Largely influenced by Skinner's observations that reinforcement schedules influence the choice of response rates, modern psychologists such as Killeen have gone back to mathematically explain 'arousal' in terms of schedules of reinforcement (Killen & Hanson, 1978). In his experiments, pigeons were presented food once a day in a feeding chamber and afterwards, were placed into an enclosure with floor panels that could measure their activities. They found that immediately after feeding, activity levels were high, and then followed an exponential decay with a fixed constant. They termed this heightened activity state 'arousal'. Next, they increased the frequency of feeding to various fixed time intervals several times a day and noticed that activity levels increased substantially. Thus, they determined that if the time between successive reinforcements are short enough, arousal levels will 'cumulate to an equilibrium level' that can be predicted from the exponential decay constants (Killeen, Hanson et al. 1978). Thus, one of their main principles of reinforcement was $A=ar$, where A, arousal, is proportional to the r, the rate of reinforcement and a is a

'constant of proportionality'. These studies showed that general arousal, as measured by the rate of activity, is proportional to the average rate of reinforcement.

Although the study of response rates and motivational drive dominated much of the mid- to late 20th century, it largely yielded to the rapid rise and great success of the study of another facet of decision making, action selection.

Models of action selection: Valuation, Selection, and Re-evaluating

Action selection is a process of choosing the best action among a set competing and often incompatible potential actions. These alternative action plans vie for control of motor resources and in a 'winner take all' fashion, an arbitrator decides which of the actions or action sequences will be allowed expression. Optimally, the winning action is a response that yields the best expected outcome, given the internal and external state of the animal.

The study of choice began as early as the 17th century, where Blaise Pascal initiated the notion of the 'expected value.' The expected value of an action is computed by multiplying the value and probability of occurrence of each potential action, where the rational choice is the action that yields the highest expected value (Pascal, 1670). Bernoulli later initiated 'expected utility theory', which was able to better explain why people's behavior often violated Pascal's theory, such as when making decisions that involve risk. For example, if given the choices A and B, where A is 95% chance at winning $1 million and B is 50% chance at winning $3 million, most people prefer A, even though expected value

theory dictates that B is valued higher. By taking into account the fact that people are risk-averse, choice behavior can be better predicted by calculating expected utility, or the amount of satisfaction that is expected from the outcome. von Neumann and Morgenstern later formally provided a set of necessary and sufficient axioms of expected utility theory that define a rational decision maker (Von Neumann and Morgenstern 1944). Later, Kahneman and Tversky showed through a series of experiments that people's actual behaviors are even less rational than what the normative expected utility models describe. They subsequently developed prospect theory, which provided psychologically more accurate descriptive models of human choice, introducing effects such as framing and reference-dependence (Tversky and Kahneman 1974). These frameworks encouraged questions in neuroscience such as how the subjective utilities of different actions are acquired, how context guides the valuation of options, and ultimately, how subjective values guide decisions.

The study of the neural mechanisms underlying valuation, or how actions are assigned subjective value, has been powerfully influenced by computational models of reinforcement learning such as Rescorla-Wagner models of reinforcement learning and later, temporal-difference learning models (TD-models). As behaviorally demonstrated by Thorndike's Law of Effect, actions followed by pleasant consequences are more likely to be performed, while actions followed by negative consequences are dampened (Herrnstein 1970). Reinforcement learning models provided a way to explain how actions leading to positive outcomes are reinforced. At the heart of reinforcement learning models is the concept of *reward prediction error*, which is the discrepancy between actual and expected outcomes. Rescorla-Wagner models originally used the discrepancy between expected and

actual outcomes to describe how the Pavlovian values of stimuli can be acquired on a trial-by-trial basis (Rescorla 1971). For example, an outcome that is better than predicted at the end of each trial can generate an error signal that teaches or updates the value of the outcome-predicting stimulus. A later extension of the Rescorla-Wagner model is the more general TD-learning model, which considers a continuous flow of time rather than dividing time into discrete trials (Barto, Sutton et al. 1983). The key to the TD model is that instead of receiving a *reward prediction error* at the end of a 'trial', one can receive a prediction error at any time *t* if one had previously made a wrong prediction at time *t-1*. This provides a model for decisions that require multiple intervening steps before an outcome is realized, which is particularly useful in the case of action selection, where oftentimes, several actions may be made before reward is realized. Further, it is key for solving the problem of credit assignment, which is figuring out which action should be attributed to the rewarding outcome.

Action values, or the expected values of different actions, can be thus learned and continually updated by reward prediction error signals. The action value Q, of action *a* is the expected reward *r*, given action *a:*

$$Q(a) = E[r|a]$$

If action values are learned for all alternative actions, then the best choice would be to choose the action that yields the highest Q(a).

Moreover, the TD model of reinforcement learning is an essential feature of actor-critic models, one of the most influential models of action selection (Sutton and Barto 1998). The actor-critic model in its earliest form was formulated by Barto and Sutton, where

two adaptive elements, an actor and a critic, interact to flexibly carry out plans of action depending on the state. In this model, a critic, using TD-learning, learns to predict expected reward given a set of states and actions. The actor uses the critic's information and maintains a policy of what actions should be carried out. Thus, the tendency to perform certain actions over others is based on the reward prediction error signals generated after performance of the action.

Theories of reinforcement have been extremely successful in that much of the essential elements of the models, reward prediction error, action value representation, and actors/critics have been experimentally linked to particular neural substrates, which will be reviewed later in this chapter.

Although reinforcement learning models are extremely powerful for studying action selection, their limitation is that standard models of reinforcement learning cannot account for how animals decide *how* to act or *whether* to act at all. To date, very few studies have made careful attempts to explain both types of selection.

Models seeking to combine motivational drive and action selection: extending standard reinforcement learning models

The success of TD-models in action selection has inspired some computational neuroscientists to extend these models to include explanations of vigor selection. McClure et

al., 2003 modeled overall running rate as a series of binary choices between running and not running (McClure, Daw et al. 2003). The model sequentially went through each binary choice and either accepts or rejects the act of running; if an action is not accepted, time is incremented forward and the next action of the series is considered. From this, they can incorporate the effects of prediction error signals on a continuously varying response vigor function. In their model, motivation is modeled as a cached, time-discounted accumulation of reward values acquired by TD learning. The elimination of prediction error signals reduces the probability of accepting actions, effectively reducing 'running speed'. Such a model is among the first to extend reinforcement learning rules on the choice of response vigor.

Niv and colleagues developed one of the first normative models of response vigor selection (Niv, Joel et al. 2006; Niv 2007). In this model, the choice of response vigor involves a different kind of optimization process from action selection. Foremost, they assume that the faster or more frequent the response, the earlier reward will be obtained. However, faster or more 'vigorous' responses are more energetically costly to the animal. The key optimization problem for response vigor selection is that animals balance a tradeoff between the energetic costs of performance with the benefits of rapid reward acquisition. One consequence that follows from the above assumptions is that the cost of moving slowly delays *all* future events. Thus, the optimization process involves animals balancing the cost of performance against the delay of *all* future rewards, or the expected *net rate of reward*.

Finally, the key to the normative model of motivation selection is that the net rate of reward has global, not proximal effects on vigor. For example, when the net reward rate is high, the opportunity cost of wasted time is high. Thus, animals should act as fast as possible for *all* actions, as not to waste precious time. Niv, Dayan, et al. determined what was the optimal choice of response rate in simulations of rats performing with varying levels of motivation (hungry versus satiated). Highly motivated simulated rats (hungry rats) not only pressed levers for food faster, but also performed all other actions such as grooming or pulling chains that release water faster. Formally, the simulated animal repeatedly chose action-latency pairs, or (a, τ), where a is the action and τ the latency, which was used to formalize 'vigor'. With each action choice, τ time passes, and no other actions are executed during that time. Further, each action incurs a τ or latency-dependent vigor cost. Finally, the goal of the animal was to optimize net utility per unit time, which is the sum of all rewards minus all costs incurred. For hungry rats, the optimal way to maximize net utility required animals to choose not only actions that yield food but also τ to be shorter for *all* action-latency pairs (Niv, Daw et al. 2007).

Intriguingly, this model provides a normative framework for describing old and controversial observations regarding generalized drive, first proposed by Hull, who noticed that hungry or sexually deprived animals seemed ‘jumpy’ or generally more active. Moreover, this provides an explanation for how the energizing effect of motivation can globally motivate behaviors, in contrast with the directing effect of motivation (or action selection), which is outcome-specific.

Models of behavior, particularly those related to reinforcement learning have powerfully guided the design of experiments in decision making. In turn, key components of these models such as prediction error signals and action value signals have been substantiated by experiments performed in rodents, non-human primates, and humans. In the following, I highlight some of the most prominent neural substrates that are thought to participate in different aspects of decision making. I start from neural substrates of action selection, which have been the most clearly characterized, and end with studies that provide hints on the neural substrates underlying the selection of motivational drive.

Neural substrates of decisions making

Action selection and the functional anatomy of the basal ganglia

The neural basis of action selection has largely focused on the basal ganglia and its corresponding cortical structures. In the following, we first describe the basic functional anatomy of the basal ganglia-cortical circuitry and then, how different components may be mapped onto different sub-processes of action selection.

Cortico-basal ganglia loops: general anatomy

The striatum is the largest and principal input nucleus of the basal ganglia, receiving excitatory input from virtually every part of the cerebral cortex and associated thalamic

regions. Cortical input maintains rough topography, but there is both convergence and divergence of inputs onto striatal neurons (Selemon and Goldman-Rakic 1985; Alexander and Crutcher 1990; Parent and Hazrati 1995). Medium spiny neurons (MSNs) of the striatum send inhibitory input, either directly or indirectly to the output nuclei of the basal ganglia, which are the globus pallidus internal segment (GPi) and the substantia nigra pars reticulata (SNr). The output structures of the basal ganglia, GPe and SNr, receive topographically preserved input from the striatum. They send inhibitory projections to the thalamus, which then project back to the cortex, thus completing the cortico-basal ganglia loop.

In addition to its reciprocal connections with cortex, the basal ganglia is innervated by and projects to dopaminergic neurons. Dopamine (DA) neuron projections are highly divergent, broadcasting to 1000 times as many targets (e.g. striatum, cortex, amygdala, etc) as there are DA neurons (Joel and Weiner 2000). The striatum receives the heaviest DA projections, which can modulate the strength of cortico-striatal synapses, potentially to weaken or amplify certain types of incoming cortical information targeting the striatum (Reynolds and Wickens 2002). Furthermore, the connections between striatum and dopamine in rats are topographically organized. The dorsolateral striatum (DLS) is reciprocally connected with the lateral SNc. The dorsomedial striatum (DMS) is innervated by medial SNc but projects to both medial and lateral SNc. Lastly, the ventral striatum (VS) is innervated by VTA but broadly projects back to VTA and SNc. Thus, the connections between DA and striatum are asymmetric, in that the VS broadly influences DA neurons that project to all three striatal regions, the DMS influences DA neurons that affect both DMS and

DLS, and the DLS's influence is most restricted (Haber, Fudge et al. 2000; Joel and Weiner 2000). This pattern, however, is less clearly delineated in primates. Yet in general, all three regions of the striatum in both rats and primates connect with cortex and dopamine neurons in a roughly organized manner.

Next, we describe how the striatum can be parsed into functionally separable modules on multiple levels.

Direct/indirect pathways

Striatal MSNs comprise over 90% of striatal neurons and are inhibitory projection neurons of the striatum. MSNs can be distinguished into D1- and D2-type dopamine receptor-expressing cells that are biochemically and morphologically distinct. These two cell types are generally distributed in a salt and pepper fashion throughout the entire striatum (Le Moine and Bloch 1995) but more D2 may be present in dorsolateral striatum (Joyce, Sapp et al. 1986). D1-expressing MSNs comprise the 'direct' pathway by projecting directly to GPi/SNr. D2-expressing MSNs, on the other hand, project indirectly to the GPi/SNr, via a detour through the globuspallidus external segment (GPe). The net effect of the direct pathway is the promotion of movement (because MSNs disinhibit GPi/SNr activity) while indirect pathway suppresses movement. These two opposing pathways that intriguing rely primarily on inhibitory signals may be involved in the focused selection of desirable movements (Gurney, Prescott et al. 2001). Recent studies showed that optogenetically

activating D1 facilitated movement while D2 suppressed movement (Kravitz, Freeze et al. 2010). Moreover, perturbing D1 and D2 activity in monkeys using antagonists showed that differential effects on saccade reaction time biases (Nakamura and Hikosaka 2006).

Patch and matrix

The striatum has additionally been divided into acetylcholine-esterase (AChE)-rich 'matrix' compartments and patches of AChE-poor 'striosomes'(Graybiel and Ragsdale 1978). Analogous to 'swiss cheese', matrices are the cheese and patch striosomes are the holes. There is some evidence that patches are more enriched in the rostral and medial poles of the striatum while matrix are relatively more prominent in the posterior and lateral regions (Desban, Kemel et al. 1993). Although both patch and matrix compartments receive input from all areas of cortex, patch compartments receive relatively dense input from periallocortical areas and relatively less from neocortical areas (Gerfen 1992; Nakano, Kayahara et al. 2000). Patch and matrix neurons also generally do not cross compartmental boundaries (Walker, Arbuthnott et al. 1993), suggesting that these compartments are functionally separable. Houk, Adams and Barto have proposed that patch and matrix architecture may be mapped onto actor/critic models in that patch striosomes form the critic while matrices form the actor, although this has yet to be proven experimentally (Houk, Adams et al. 1995).

Parallel loops: associative, sensorimotor, and limbic

Finally, the cortico-basal ganglia circuit is comprised of topographically preserved loops, suggesting that they be used for parallel and distributed processing of information. In support of this, cortico-basal ganglia loops are divided into three main anatomically and functionally segregated loops: sensorimotor, associative, and limbic loops (Middleton and Strick 2000).

'Sensorimotor loops' are formed between somatosensory and motor cortical areas and the DLS. Projections are somatotopically preserved throughout the loop, where DLS projects to the ventrolateral portion of the GPe and GPi, which project to the ventral lateral portions of thalamus, which finally feeds back into sensorimotor regions. The associative loops are formed between the DMS and prefrontal associative cortical areas such as medial prefrontal, anterior cingulate, orbitofrontal, and dorsolateral prefrontal areas (Uylings et al., 2003). DMS striatum projects to medial portions of the GPi, which projects to medial thalamic nuclei, which feedback to medial frontal cortical regions. Lastly, the limbic loops connect VS, comprising nucleus accumbens and olfactory tubercle, with hippocampus, amygdala, entorhinal cortex, perirhinal cortex, medial orbitofrontal cortex, and parts of the anterior cingulate cortex (Haber, Fudge et al. 2000; Bornstein and Daw 2011).

Several lesion studies have supported the functional segregation of associative, sensorimotor, and limbic loops. Studies by Yin, Knowlton and Balleine showed that DMS (associative striatum) lesions impaired flexible, goal-directed but not habitual behavior (Yin, Knowlton et al. 2005; Yin, Ostlund et al. 2005). Along similar lines, Killcross and Coutureau (2003) showed that lesions of the prelimbic associative cortex (also part of the

associative loop) caused otherwise goal-directed behavior to become habitual (Balleine and Dickinson 1998; Killcross and Coutureau 2003). On the contrary, lesions of the DLS rendered animals permanently goal-directed and prevented the formation of habits (Yin, Knowlton et al. 2004; Yin, Knowlton et al. 2006). Thus, associative loops may be involved in flexible, goal-oriented behavior, while motor loops are necessary for the formation of habitual behaviors. Lastly, structures of the limbic loop such as VS and the amygdala are implicated in stimulus-reward associations, Pavlovian prediction learning, motivation, and addiction (Mogenson, Takigawa et al. 1979; Ikemoto and Panksepp 1999; Hall, Parkinson et al. 2001; Cardinal, Parkinson et al. 2002; Salamone and Correa 2002). Moreover, these three subregions of the striatum have been hypothesized to map onto different subcomponents of actor-critic models of action selection. Supported by fMRI studies in humans, the VS may function as a critic while the dorsal striatum (DS) is the actor (O'Doherty, Dayan et al. 2004). Moreover, the DMS may be the actor for model-based selection (where animals rely on knowledge of the dynamics of the environment) while the DLS is for model-free selection (Bornstein and Daw 2011).

In sum, the basal ganglia circuit can be organized into parallel channels, making it a potential site to process multiple and possibly competing types of information (cognitive, motor, affective). It remains mysterious what functions the various loops of the basal ganglia perform, how separable these loops are, and how they interact or compete with one another. Thus far, there are a growing number of studies teasing apart different parts of the circuit and mapping them onto different aspects of decision making. To examine whether and how different components of the circuitry participates in action selection, electrophysiological and

lesion studies were used to measure and perturb the activity of specific parts of the circuit, respectively.

Mapping neural substrates to components of action selection and motivational drive

Phasic DA activity encode reward prediction error

Activity of dopamine (DA) neurons recorded by Schultz and colleagues provided the most famous and compelling support for Sutton and Barto's reinforcement learning models (Schultz 1998). Phasic bursts of dopamine neurons encode reward prediction error, which is the discrepancy between the value of predicted and actual outcomes. Therefore, DA neuron phasic activity was mapped to the teaching signal critical to TD models of reinforcement learning. This discovery inspired the idea that DA neuron activity may be used as a teaching signal for DA neuron's projection targets such as striatum, which can flexibly learn the values of stimuli and actions. The vast amount of literature supporting DA's role in encoding reward prediction error has largely solidified the study of action selection and motivation around structures closely connected to DA, particularly the striatum.

Value representation in the dorsal striatum

Kawagoe et al., (1998) recorded from caudate neurons in monkeys performing a memory-guided saccade task, where monkeys were instructed to move their eyes in different

directions. The value of the locations varied, as certain directions were rewarded or not rewarded. They discovered that during the delay period between cue and movement, single neurons in the caudate were modulated (by exhibiting higher or lower firing rate) by the expected reward value, providing among the first demonstrations of neural activity linking actions to reward (Kawagoe, Takikawa et al. 1998). Furthermore, Lauwereyns et al., (2002) varied reward expectation associated to specific saccadic directions systematically in blocks of trials and demonstrated that caudate neurons show biased activity for (contralateral) high value directions even *before* the onset of a cue instruction (Lauwereyns, Watanabe et al. 2002). These studies suggested the caudate's role in assigning value to action as well as its possible influence on action selection.

In a study by Samejima and colleagues, monkeys were required to choose between one of two actions, where the reward probabilities of these actions were systematically varied in blocks of trials (Samejima, Ueda et al. 2005). Recordings showed that a large fraction of neurons in the DS varied with the values of actions even before monkeys revealed their choice, suggesting that they encoded 'action values'. Their task allowed them to dissociate 'action value' from the actual action the monkey subsequently made, as these action value activities were independent of what the action the monkey later performed. Importantly, action value neurons encoded the value of specifically one choice or the other, but not both, indicating that they track the values of specific choices in an absolute manner, or a 'menu-invariant' way. This result was later supported by Lau and Glimcher, who found similarly large proportions of action value neurons in the dorsal striatum (~30% of recorded neurons) prior to choice (Lau and Glimcher 2008). In total, these studies suggested that the values of

potential actions are represented in an absolute manner in the striatum, allowing mapping of striatal neuron activity to the valuation process of decision making.

In further support of the DS' role in action value representation, other studies compared action value representation in the striatum with that of its upstream or downstream structures. Pasquereau et al., 2007 compared activity in DS with neurons in downstream GPi. While action value neurons were found in both nuclei, there were significantly more 'chosen value' neurons in the GPi, which are neurons encoding the value of the already 'chosen' action (chosen value signals were also found in the OFC by Padoa-Schioppa and Assad, 2006, as well as in DS by Lau and Glimcher, 2008). Moreover, Seo et al., 2012 compared DS neurons with its associated cortical structure, the lateral prefrontal cortex (lPFC) and found that the DS is more enriched in action values representation than the lPFC (Seo, Lee et al. 2012). Thus, it has been suggested that action value representations may arise first in the striatum, whereas selection may be performed somewhere downstream (Daw and Doya 2006).

It is worthwhile to note that despite a convergence of evidence in support of the monkey DS's role in encoding action value, much fewer action value neurons were found in rats before the manifestation of choice (Kim, Sul et al. 2009). Yet, Kim et al., did find significant proportions of chosen value coding neurons that manifest after choice was already made, which are consistent with previous studies in monkeys (Lau and Glimcher, 2008). Finding few action values before the expression of choice but significant proportions of chosen value signals suggested that the DS might be more involved in evaluation rather than

selection itself. Lastly, fMRI studies showed after scanning through the entire human brain, action value signals before choice was most prominent in the human supplementary motor cortex instead of DS (Wunderlich, Rangel et al. 2009).

Moreover, recent studies found that DS also encodes the relative values of actions (neurons were significantly modulated by the *difference* between the values of two actions)(Cai, Kim et al. 2011). This study suggests that DS may not be restricted to representing the values of actions in an absolute manner, and may be more directly involved in the selection process than previously thought.

Lastly, it is worth mentioning that the question of whether selection is based on action values or values of more abstract entities such as 'goods' is also under debate (Padoa-Schioppa and Assad 2006; Wunderlich, Rangel et al. 2010; Padoa-Schioppa 2011). In action selection, choices are 'embedded in pre-motor processes', meaning that the values of goods or stimuli are passed onto the values of actions, and selection takes place in 'action value space'. Thus far, there has been evidence of the absolute representations of 'goods', unrelated to visual or movement signals in the primate orbitofrontal cortex (OFC) (Padoa-Schioppa and Assad 2006). Wunderlich et al., in an fMRI study has also provided support for selection based on 'goods' in humans. In their task design, subjects chose between different stimuli, with or without knowing what actions are needed to obtain them. They found that the ventromedial cortex encodes the value of the stimulus the subject will choose, even before choice is revealed and irrespective of the subject's actions. Thus, some types of

selection may take place in more abstract space. It is likely that both types of selection may take place in the brain.

Relative value coding in cortex: Lateral intraparietal cortex, anterior cingulate cortex, medial prefrontal cortex

In addition to representing the values of potential actions using a common neural currency, the values of different actions must be compared relative to one another for appropriate selection to occur. Relative value signals depend on the values of both action values (in the case of two choice tasks), and correlate with the animal's subsequent choice. A well-known area thought to encode the relative values of actions is the lateral intraparietal cortex, or LIP. The LIP is a higher-order visual area that promotes visually-guided behaviors and attention such as saccadic eye movements. LIP neurons appear to encode the value of an action (a saccade into the neuron's receptive field), relative to other possible actions (saccades outside the neuron's receptive field) such that the more valued action has higher firing rate and thus, a higher chance of being selected for motor execution (Platt and Glimcher 1999; Platt 2002; Dorris and Glimcher 2004; Sugrue, Corrado et al. 2005; Seo, Barraclough et al. 2009). Moreover, these results may be interpreted to support the notion that alternative action plans can be embedded in motor structures and that action selection can be performed in 'action space'.

Other cortical areas such as anterior cingulate cortex (ACC) and dorsomedial prefrontal cortex (mPFC) have been also implicated in the selection process. Single neuron

recording in the anterior cingulate cortex of monkeys and in the mPFC of rodents, has been shown to encode the difference between action values (Seo and Lee 2007; Sul, Kim et al. 2010). Along similar lines, fMRI studies in humans have found relative value signals in both the ACC and dmPFC (Wunderlich, Rangel et al. 2009), which interestingly reflected the value of the *non*-chosen value minus the chosen value (as opposed to the more intuitive *chosen* minus non-chosen value).

Neural substrates of response vigor selection

As reviewed above, the theoretical basis for response vigor selection has paled in comparison to action selection. In the following, we briefly discuss lesion literature that implicates certain striatal regions to be involved in motivation. Finally, we introduce some studies that may provide the hints about what brain regions represent net reward value, which Niv, Dayan, and colleague's model predicts to be critical in response vigor selection.

A wide range of studies have tied dopamine levels in the striatum, especially the nucleus accumbens to the invigoration of responding (Berridge and Robinson 1998; Ikemoto and Panksepp 1999; Salamone, Correa et al. 2007). Several lines of work from Salamone's lab indicated that nucleus accumbens is required for overcoming effortful work demands. Rats were trained on reinforcement schedules requiring varying levels of effort. For example, FR1, FR4, FR16, are schedules requiring increasing effort. Dopamine depletion in the NAc had no effect on the lower effort schedule, but it substantially reduced responding on the task requiring higher demand (Aberman and Salamone 1999; Salamone, Wisniecki et al.

2001). In another study from the same group, dopamine depletion of the NAc deterred rats from choosing actions that required more effort yet high reward, in favor of actions that required less effort and lower reward (Cousins, Atherton et al. 1996). These results suggest a role for dopamine in the VS for balancing cost-benefit tradeoffs.

A role for balancing energetic costs with reward benefits have also been found in the ACC through lesion studies of rats performing similar tasks that required choices between high effort/high reward and low effort/low reward (Walton, Bannerman et al. 2003; Walton, Kennerley et al. 2006). Further, human fMRI studies have pointed to the ACC and VS in effort-based cost/benefit evaluation (Croxson, Walton et al. 2009).

Although many past studies have converged onto the VS's critical role in motivation, there is evidence for the dorsal striatum (DS)'s role in motivation as well. Human fMRI studies that compared hemodynamic responses to rewards and punishment found that the intensity of signal in the head of the caudate (anterior DS) was higher than in VS (Delgado, Stenger et al. 2004). Human PET studies have shown that DA levels in the DS increase for non-hedonic food motivation (presentation of food without consumption) (Volkow, Wang et al. 2002). Further, although the VS is famously involved in mediating the effects of drugs of addiction (Everitt and Robbins 2005), it's been observed that cues that predict drugs strongly engage the DS in humans (Volkow, Wang et al. ; Wong, Kuwabara et al. 2006). Further, a series of viral restoration and pharmacological studies using in mice showed that the nigrostriatal pathway from SNc to DS was necessary and sufficient for initiating goal-directed behaviors (Palmiter 2008). In rats, dopamine levels increased in the DS in response

to drug-associated cues (Ito, Dalley et al. 2002), while dopamine antagonists in DS weakened the effects of cue-elicited drug seeking behavior (Vanderschuren, Di Ciano et al. 2005).

Net or 'state' value coding neurons in the ventral striatum, ventral pallidum, ACC, LIP, and dorsal striatum

Finally, as several past theories of 'arousal' and vigor selection predicts the importance of tracking net reward value, we describe work that has uncovered such neural responses, even if they do not explicitly link net value to the selection of motivational vigor. In rats, neurons encoding the sum of action values (called 'state value' coding neurons) have been found in VS as well as its downstream target, the ventral pallidum (Ito and Doya 2009). Similarly, state value coding neurons have also been found in the monkey VS (Cai, Kim et al. 2011), LIP (Seo, Barraclough et al. 2009) and ACC (Seo and Lee 2007). In support of findings in non-human primates, Kolling et al. examined foraging type behavior in humans and found that the BOLD activity in human ACC varied with average reward value (Kolling, Behrens et al. 2012). Lastly, recent studies showed that neural firing rate in DS reflected changes in response vigor in monkeys (Opris, Lebedev et al. 2011) and Cai et al. unveiled a previously unnoticed but significant proportion of DS neurons that track the sum of action values (Cai, Kim et al. 2011).

CHAPTER 1

Teasing apart the action selection and the selection of motivational drive

Introduction

Motivation has been hypothesized to have an energizing and directing component (Dickinson and Balleine 1994; Salamone and Correa 2002; Niv 2007). While the directing component steers behavior towards specific actions, the energizing component is hypothesized to globally motivate behavior. Current studies of decision making have largely focused on the directing aspect, or the problem of action selection. In most decision making paradigms, animals are trained to select among a discrete set of actions in pre-set time windows, often keeping inter-trial intervals under the control of the experimenter. Thus, animals often do not choose their own rate of performing trials, often obscuring the animal's level of engagement in the task. Yet, it's evident that animals appropriately allocate their responses over time according to the structure and demands of their environment (Ferster and Skinner 1957). Therefore, we aimed to design a task that probes not only how animals choose actions but also how they decide *whether* or *to what extent* they engage in the task.

Currently, it is still mysterious which brain regions are critical for allowing one to flexibly select levels of motivational drive. To date, there is evidence to support the striatum's role in motivation. The ventral striatum (VS) has been historically tied to motivation and was famously alluded to as the 'interface between motivation and action' (Mogenson, Jones et al. 1980; Kelley 2004). This view is supported by a massive amount of literature linking the VS to the motivational effects of drugs of addiction (Everitt and Robbins 2005), balancing the costs of performance with the benefits of reward (Cousins, Atherton et al. 1996; Salamone, Correa et al. 2007), and encoding state value (Ito and Doya 2009; Cai, Kim et al. 2011).

In addition, there is emerging support for the dorsal striatum role in motivation in rodents (Szczypka, Kwok et al. 2001; Sotak, Hnasko et al. 2005; Robinson, Rainwater et al. 2007; Palmiter 2008), non-human primates (Nakamura and Hikosaka, 2006), and humans (Delgado et al., 2004; Volkow et al., 2002,). In particular, lesion studies have found that the dorsomedial portion (DMS) is most critical for promoting flexible, goal-oriented behavior (Yin, Knowlton et al. 2005; Yin, Ostlund et al. 2005) , for mediating reaction time and initiation latency (Brown and Robbins 1989; Hauber and Schmidt 1994; Bailey and Mair 2006), and reversal learning (Ragozzino 2007). Finally, several electrophysiology studies showed that DMS neurons flexibly encode stimulus-values (Kimchi et al, 2008), correlate with learning (Thorn et al., 2010), and encode the sum of action values (Cai, Kim et al. 2011).

In all, the DMS and VS both have potentially important roles in energizing motivation. Here, we designed a task that systematically examines the energizing and directing aspects of motivation in a unified task. After dissociating these orthogonal processes, we performed selective lesions of the DMS and VS to examine their differential effects on the two aspects of decision making.

Results

Self-paced decision task

To study the energizing and directing aspects of motivation, we designed a self-initiated, two-alternative choice task. Rats initiated each trial by poking their snout into a central odor port. Upon poking into the odor port (trial initiation, or 'odor-poke-in'), a randomly selected odor cue was presented after a variable (300-500ms) delay, which signaled for the animal to poke into either a left or right port for a water reward (**Fig.1.1a-b**)(Uchida and Mainen 2003). To parametrically manipulate the difficulty of the task, we used pure odors as well as binary mixtures of odors of various ratios, where the dominant component in the mixture instructed the animals which side will be rewarded. For example, 100% or 60% odor A signaled the animal to go to the left side for reward while 100% or 60% odor B signaled the animal to go right. Incorrect choices were unrewarded and occurred primarily for the mixed-odor trials, which are more difficult to discriminate.

Concurrently, to examine the effects of reward values on behavior, we systematically varied the amount of water delivered at the left and right ports in blocks of 40-60 trials (**Fig.1c**, blocks are color-coded) (Samejima, Ueda et al. 2005; Rorie, Gao et al. 2010). For example, reward delivered at the left water port is the same in both cyan and red blocks. However, the left reward port yields relatively more water than the right port in cyan blocks, while the left reward port yields relatively less water in red blocks. On the other hand, while the net reward rate is higher in red blocks, the net reward rate is lower in cyan blocks.

The relative value of potential actions directs choice biases

We examined how the values of the options affected behavior. First, consistent with previous studies, (Samejima, Ueda et al. 2005; Rorie, Gao et al. 2010) the animal's choices were biased by the relative value between the two reward ports (**Fig. 1.1d** and **Fig. 1.1f)**. In the blocks where the right port delivered relatively more water (**Fig. 1.1d** red and pink blocks), the animals chose the right side more frequently (**Fig. 1.1d** shifts in psychometric curves towards the right reward port in red and pink). Moreover, this bias towards the relatively more valuable side did not depend on the net value of the block (red and pink curves are not significantly different). Thus, the relative values of the options, but not net value, affected the choice biases of the animal.

Latency to trial initiation reflects the net value of potential actions

In addition to measuring how relative values of choices bias action selection, our self-initiated task allowed us to examine the rats' motivation to perform the task, or their response vigor. To do this, we measured the amount of time it takes rats to initiate a new trial after the completion of a previous trial. Opposite to the effects of relative values, animals displayed shorter trial initiation latencies in blocks of high net value (red and blue) compared with blocks of low net value (pink and cyan) (**Fig. 1.1e:** we obtained the median trial latency time per block type per animal and computed the average and standard error. Bar graphs are color-coded by block type). Furthermore, alternations in trial initiation time across changes in net value are reflected within 10 trials of a block change (**Figure 1.1g**).

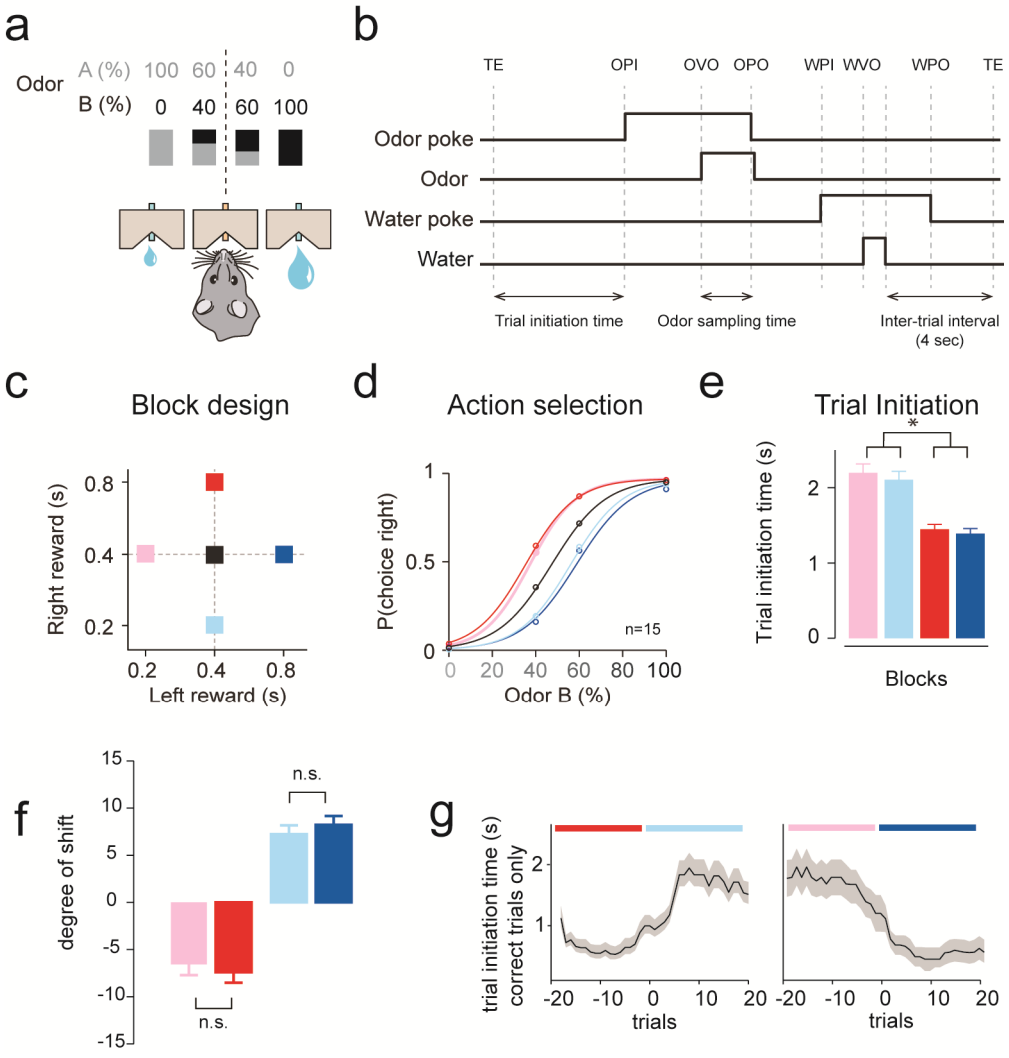

Figure 1.1 A self-paced decision making task: dissociating effects of relative and net value on choice and motivation

(a) Odor mixture categorization task. **(b)** Task epochs. TE, trial end; OPI, odor poke in (entry);OVO, odor valve on (opening); OPO, odor poke out (exit); WPI, water poke in; WVO, water valve on (opening); WPO, water poke out (exit). **(c)** Reward values at the left and right ports varied across blocks. Each colored square represents a block. (d) Psychometric choice curves in different blocks. Colors refer to the blocks in **c**. **(e)** Trial initiation time (mean ±s.e.m.). **P*<0.05, T-test. **(f)** Degree of shift across blocks. **(g)** Time-course of trial initiation time across blocks.

The energizing effect of net value is global and independent of previous trial outcome

The key feature that distinguishes the energizing from the directing aspect of motivation is that the energizing aspect affects actions globally. However, our results on trial initiation time presented thus far can still be sufficiently explained by the *directing* effects of motivation alone, without the need to introduce the variable 'net value': rats may appear slower in blocks of 'low net value' under our task design, but may simply be slower to initiate after receiving small rewards (which occurs more frequently in low net value blocks). Conversely, they may be faster to initiate trials after all large rewards, which coincides with blocks of high net value. If the overall net value of the block is the main energizer of trial initiation time, initiation time should depend primarily on the net value of the block rather than on the previous trial's outcome. Thus, we examined trial initiation time after different trial outcomes to see whether trial initiation time is truly net-value dependent or if it can be explained by other factors.

In **Figure 1.2b,** we show boxplots and cumulative distribution function plots of trial initiation time after either correct trials or error trials across different block types. Regardless of correct or error outcome, animals were faster in high net value blocks and slower in low net value blocks. Next, we examined trial initiation time following left and right reward trials independently. We hypothesized three possibilities: trial initiation time could depend on 1) the absolute reward amount from the previous trial 2) the relative value of the previous trial's reward or 3) the net value of the block. For example, in the following we examine the distribution of initiation time following all left-reward trials in *cyan* and *red* blocks. The

absolute value of the left side in both blocks are identical while the relative value of left choice is higher in cyan blocks (**Fig. 1.2a** middle panel, block-wise changes in left value). If the motivation to initiate trials is driven by *absolute* values, the initiation time following left reward trials should be the *same* in both red and cyan blocks. However, if initiation time is driven by the *relatively* larger side, then initiation should be *faster* in *cyan* blocks. In fact, we observed that neither is the case. Animals were significantly slower after left reward outcomes in cyan blocks than in red blocks (**Fig. 1.2c** red versus cyan; Kolmogorov-Smirnov test, $P<0.01$). This is because the net value is lower in cyan blocks than in red blocks. Thus, we determined that trial initiation time depends on the *net value* of the block. Moreover, latency after left trials in cyan blocks was no faster than in pink blocks, where left value is smaller (**Fig. 1.2b** left trials; **Fig. 1.2a** middle panel, left reward smallest in pink) (Kolmogorov-Smirnov test, $P>0.05$). Conversely, initiation time following left trialsin red blocks was no slower than in blue blocks, where left value is even larger (**Fig. 1.2b** left trials; **Fig. 1.2a** middle panel, left reward largest in blue blocks). Lastly, we found similar patterns for initiation latencies following right reward trials as well (**Fig. 1.2c** right reward trials).

In sum, our results show that net value globally energizes task performance. Animals choose their initiation latencies according to the net-value of the block by integrating the value of both left and right reward ports, rather than responding to the absolute or relative reward size of the previous trial.

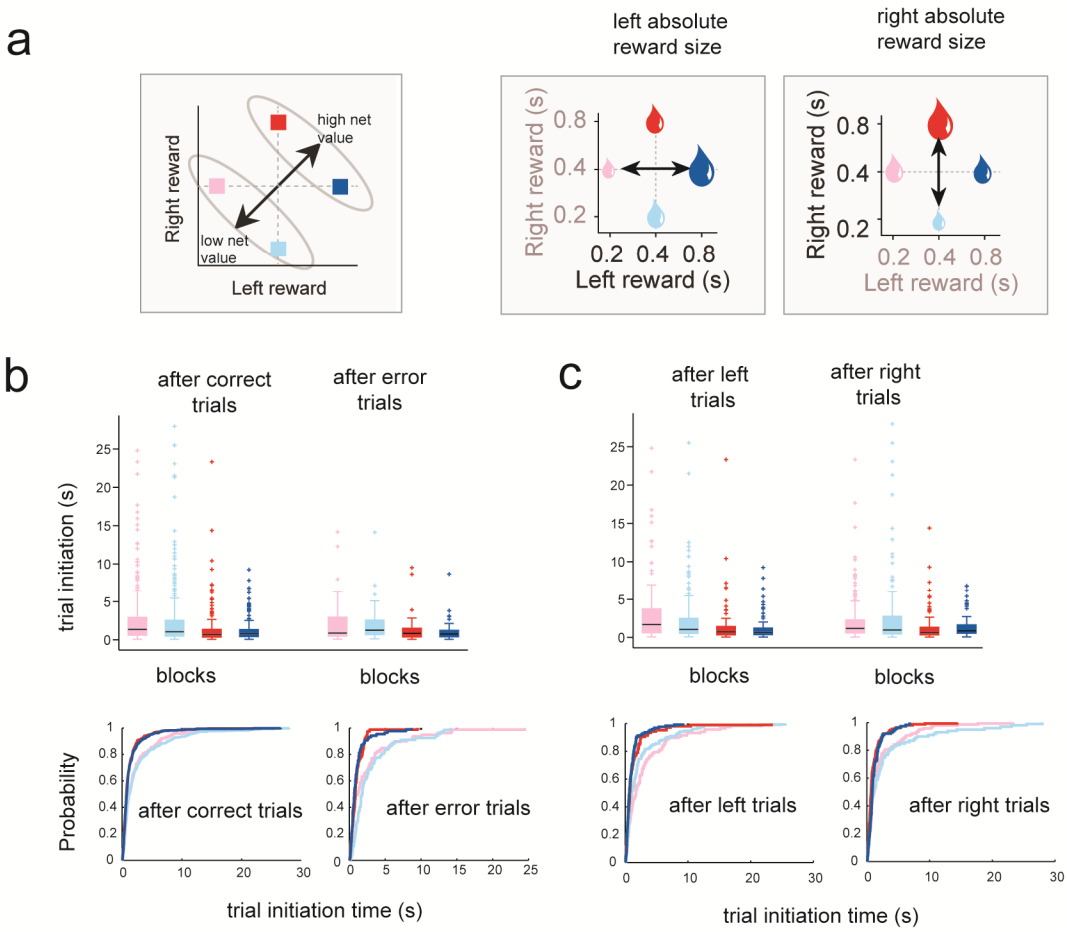

Figure 1.2 Net value globally energizes the motivation to perform
(a) Block-wise manipulation of left and right reward value. Middle and right panels: depiction of how the absolute value of left and right reward varies across blocks, respectively. **(b)** Trial initiation time after correct or error trials across blocks. Boxplots and cumulative distribution function plots are color-coded by block type. **(c)** Trial initiation time after left or right reward trials across blocks. Plots are color-coded by block type.

Past studies have theorized that motivation can be both energizing and directing, and that these two processes are orthogonal to one another (Hull 1943; Hebb 1955; Niv 2007). Here, we graphically represent the two aspects of decision making as orthogonal to one another in "value space" (**Figure 1.3a**): relative-value affect the direction of choice biases but not response vigor and thus 'directs' behaviors towards specific choices; net-value affect response vigor but not choice direction and globally energizes motivation.

Vector representation of behavior

To more systematically represent changes in motivation across blocks (i.e., how initiation latency varies with changes in left and right action value), we represented trial initiation time with a single vector projected onto "value space"(**Fig. 1.3b**) with a polar angle (θ) and amplitude (r). We set our value space in polar rather than Cartesian coordinates. This way, left value (or left 'action value', Q_L), which varies across the 'X-axis' in Cartesian coordinates, increases in the 0^o direction and decreases towards 180^o in polar coordinates. Similarly, right value (Q_R), which varies across the 'Y-axis' in Cartesian coordinates, increases towards 90^o and decreases towards 270^o in the polar coordinate. To obtain trial initiation vectors, we regressed trial initiation time with the left and right action values of each block (Q_L and Q_R, respectively) and obtained a regression coefficient for each variable.

$$\textit{Trial initiation time (s)} = b_0 + b_L \cdot Q_L + b_R \cdot Q_R \qquad (1)$$

These regression coefficients allowed us to project a single vector onto value space. We computed the square root of the sum of the squares of the coefficients to obtain the amplitude,

r. To obtain the angle, θ, we took the inverse tangent of the coefficients. The vector conveys two variables about the animal's trial initiation time (**Fig. 1.3b, left panel**): θ (in polar coordinates) reveals how trial initiation time varies across blocks (i.e. how initiation time changes with Q_L and Q_R) and *r* is the strength of the modulation (**Fig. 1.3a, right panel**). We divided the polar plot into 8 sectors and categorized vectors according to their polar angles to determine if initiation time depended on absolute, relative, or net value.

For example, if initiation latency negatively correlates with left value only (the higher the value, the shorter the initiation latency), the vector would be horizontal, or $\theta = 180^o \pm 22.5^o$ (**Fig. 1.3c**, left panel). If trial initiation negatively correlates with right value only, the vector would be vertical, or $\theta = 270^o \pm 22.5^o$ (**Fig. 1.3c**). Moreover, if trial initiation time depended on relative value, θ would be $135^o \pm 22.5^o$ ($Q_R > Q_L$) or $315^o \pm 22.5^o$ ($Q_L > Q_R$). Last, initiation latency negatively correlated with net value will be projected as a vector directed towards the lower-left-hand quadrant of value space, or $\theta = 225^o \pm 22.5^o$ (**Fig. 1.3c,** right panel). For more stringent classification, we used a 95% confidence interval on θ calculated by bootstrapping.

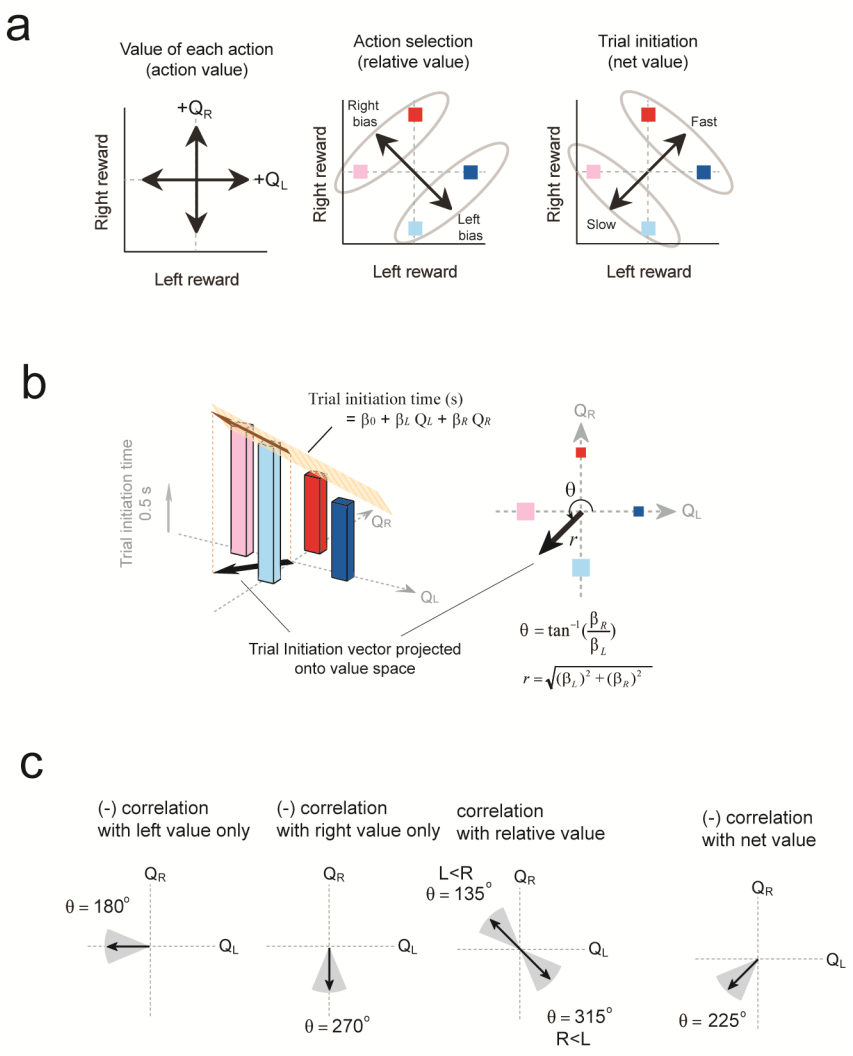

Figure 1.3 Vector representation of trial initiation time
(a) "Value space" representation. By plotting left value on the horizontal axis and right value on the vertical axis (left panel), the directing effect of relative values (middle panel) and the energizing effect of net values (right panel) can be represented orthogonal to one another in value space. **(b)** Projection of trial initiation vector onto value space. **(c)** Categorization of trial initiation vector based on the polar angle of the vector: trial initiation time can be correlated with net, absolute, or relative value.

We demonstrate our analysis on a single rat and obtain a vector of $\theta = 222^o$ (95% confidence interval: lower bound = 213^o, upper bound= 231^o) and $r = 3.3$ (**Fig. 1.4a left panel**, $P<0.01$, F-test). Since θ is near 225^o (falls within 22.5^o of 225^o) and the confidence interval did not cross 180^o or 270^o, the rat's trial initiation latency appears to be net value-dependent. The amplitude r indicates that there is an initiation time change of 3.3 seconds per second change in reward size. We then performed the same analysis on trial initiation latencies immediately following correct or error trials and obtained vectors that both indicated net-value modulation (**Fig. 1.4a left and right panels** correct trials: $\theta = 222^o$ [lower bound: 212^o, upper bound: 231^o], error trials: $\theta = 216^o$ [lower bound 192^o: upper bound 235^o]). Next, we obtained vectors for trial initiation time after left reward trials alone or right reward trials alone. Again, both vectors were net-value dependent (**Fig. 1.4b**, left panel, $\theta_L = 208^o$ [lower bound=192^o, upper bound= 223^o]; $\theta_R = 236^o$ [lower bound= 222^o, upper bound= 247^o]).

We applied the above analyses to a population of rats and found that on average, trial initiation time was net value-dependent regardless of trial outcome (**Fig. 1.4c-d,** correct trials: $\theta = 223^o \pm 2.3$, $r = 2.3 \pm 0.3$; error trials: $\theta = 222^o \pm 21$, $r = 2.4 \pm 0.4$; left trials $\theta = 212^o \pm 6.2$, $r = 2.4 \pm 0.3$; right trials $\theta = 233^o \pm 4.9$, $r = 2.5 \pm 0.3$; mean ± s.e.m.). However, we observed that θ_R (angle of vectors representing initiation time immediately after right reward) had a tendency to be larger than θ_L (angle of vectors representing initiation time after left reward) (**Fig. 1.5b,** T-test, $P<0.05$). This suggested that although initiation time primarily depends on the net value of the block (as most vectors were near 225^o), the immediately previous trial may still have an influence on initiation time.

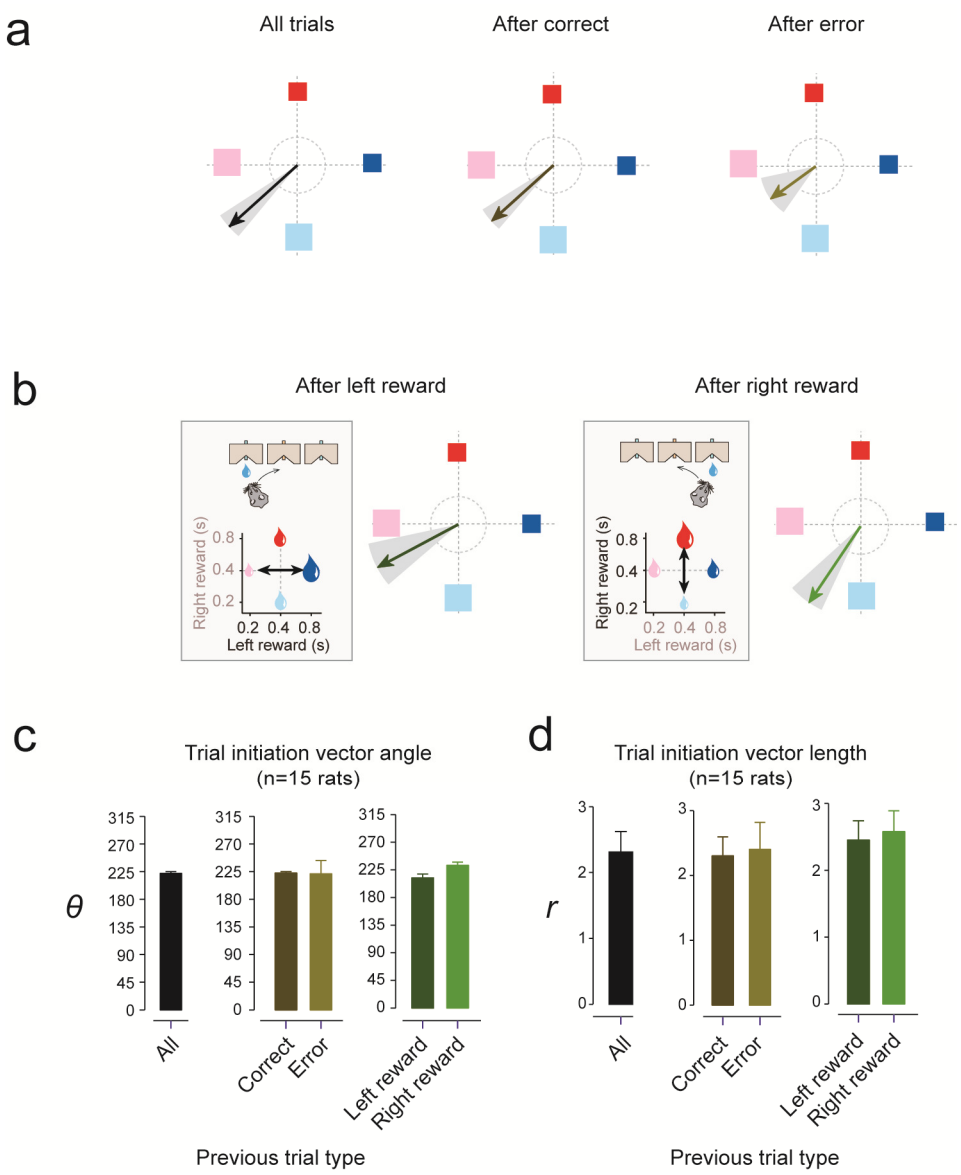

Figure 1.4 Vector representation of trial initiation time in a single rat: trial initiation depends on the net value of the block
(a) Trial initiation time after all trials, after correct only, and error only trials.
(b) Trial initiation time after left reward trials and right reward trials. Grey shading indicates the 95% confidence interval of the vector angle.
(c) Average polar vectors across rats (mean ±s.e.m.).
(d) Average polar amplitudes across rats (mean ±s.e.m.).

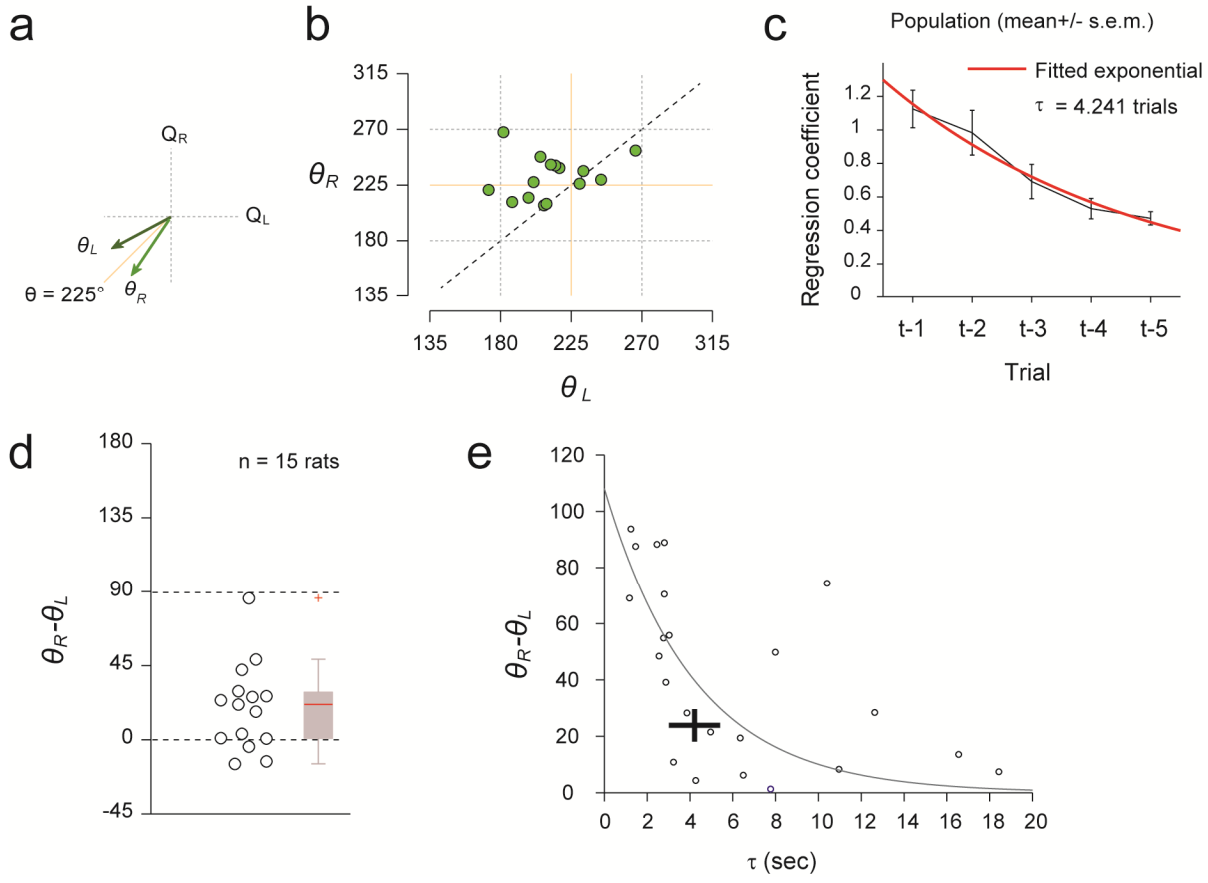

Figure 1.5 General motivation depends on the integration of outcomes of multiple previous trials

(a) Schematic showing how θ_R-θ_L is computed.

(b) Scatter plot of θ_R versus θ_L.

(c) Effect of multiple previous trial outcomes on trial initiation time. Black: regression coefficients (n=15, mean ± s.e.m.), Red: fitted exponential curve.

(d) Difference in angle between vectors representing trial initiation time immediately after left reward and after right reward trials, (θ_R-θ_L). The lower dotted line represents the median angle of trial-shuffled controls, which is the expected angle when the two vectors did not differ but has finite sample size. The upper dotted line represents the expected difference in angle if initiation time depended only on previous reward size.

(e) Simulation to determine the relationship between tau and θ_R-θ_L.
Black: simulation, Blue: fitted exponential curve. Black cross: actual tau and θ_R-θ_L(mean ± s.e.m.).

To more explicitly examine how the outcomes of multiple previous trials affect trial initiation time, we regressed trial initiation time by previous reward outcomes, where T is the current trial (e.g., Q_{T-1} is the reward outcome of one trial before):

$$\textit{Trial initiation time (s)} = \beta_0 + \beta_{T-1} \cdot Q_{T-1} + \beta_{T-2} \cdot Q_{T-2} + \ldots \beta_{T-5} \cdot Q_{T-5} \qquad (2)$$

The regression coefficients indicate the effects of past outcomes on trial initiation time. In **Figure 1.5c**, we plotted the mean regression coefficients for five past trials for the population and fitted them to an exponential to obtain a decay constant, τ, of 4.2 trials. This shows that initiation time depends on integration over multiple previous trials (average τ for 15 animals=4.6±1.5, mean±s.e.m.). We next obtained the difference in angles of trial initiation vectors immediately after left and right reward trials (θ_R -θ_L) and found that the θ_R -θ_L was very small (median: 21.3º degrees, mean: 19.6º), and was not significantly different from the θ_R -θ_L for trial shuffled controls (P>0.05, bootstrap test). This small difference in θ_R and θ_L suggested that the effect of previous trial outcome is small (**Fig. 1.5d**).

A small difference in θ_R and θ_L corresponds to less dependence of trial initiation time on immediately preceding trials and should correspond to larger τ. Conversely, a large difference between θ_R and θ_L (~90º) should correspond to smaller τ. To understand the quantitative relationship between τ and θ_R-θ_L, we performed the following simulation. Using a given value of τ, we predicted trial initiation times based on rat's actual reward history. We then obtained θ_R and θ_L, using the predicted trial initiation times for right reward and left reward trials (**Fig. 1.5e**, blue: simulated data, magenta: fitted exponential curve). This

simulation predicted that there is indeed a negative relationship between τ and θ_R-θ_L, and that θ_R - θ_L and τ obtained from the data fall within this prediction (**Fig. 1.5e**, black cross: τ and θ_{R-L}, mean and s.e.m.). This result demonstrates that our observations using trial initiation vectors and time constants (regression with reward history) are quantitatively consistent.

In all, our results show that net value energizes trial initiation for all trial outcomes, regardless of previous error or the absolute reward amount of the previous trial. In other words, the energizing drive is general rather than outcome specific. We suggest that the animals' strategy is to exploit the task during blocks of high net value because the opportunity cost of wasted time is high. In low value blocks, on the other hand, animals can afford to take their time and not respond as quickly. Furthermore, the net value affects response vigor but not choice biases. These results suggest that the two processes may be dissociable.

Selective Lesions of the Striatum

The above behavioral paradigm allowed us to simultaneously but independently measure the orthogonal effects of relative and net value on choice biases and response vigor, respectively. We next tested whether these two types of behavioral selection take place in distinct parts of the striatal circuit. Specifically, we examined what role the DMS and VS may play in either process.

We performed selective, bilateral lesions of the DMS or VS using an excitotoxic agent, ibotenic acid, after training of rats. Lesions were approximately ~2mm^2 in diameter (**Fig. 1.6a**). After recovery, we examined the choice behavior and trial initiation time of VS, DMS, and sham-lesioned animals (n=5 rats per condition; 3365.6 ± 115 trials in 7.86 ± 0.09 sessions per animal; mean ± S.D.).

Lesions of VS increased errors toward the better choice

We obtained psychometric curves for all VS, DMS, and sham-lesioned animals across different block types. None of the animals showed impaired choice biases in that all animals significantly biased their choices towards the relatively larger option of each block (**Fig. 1.6b.** dotted lines: before lesion; solid lines: after lesion), and the slopes and degree of shifts in the psychometric curves were not statistically different across conditions (**Fig. 1.6c-d** $P>0.05$, T-test). Interestingly, VS-lesioned animals had significantly more errors towards the higher valued option only for difficult, mixed-odor trials (**Fig. 1.6e,** sham: 37.9%, DMS: 38.6%, VS: 45.1% error trials towards larger choice for mixed-odor trials; $P<0.05$, T-

test). We examined whether the increase in errors or the slight trend to over-bias was due to inadequate odor sampling duration by plotting the odor sampling duration for mixed versus pure odor trials for all conditions. However, odor sampling time was not significantly different for VS-lesioned animals for either odor types (**Figure 1.6f**). Although we cannot rule out the possibility that VS-lesioned animals had impaired odor discrimination for mixed-odors, our results may support previous studies that linked VS lesions or hypo-activity to impulsivity. The tendency to make more errors towards the better option may be an indicator of such impulsivity, in that animals cannot appropriately control pre-potent actions (Cardinal, Pennicott et al. 2001; Reuter, Raedler et al. 2005; Scheres, Milham et al. 2007).

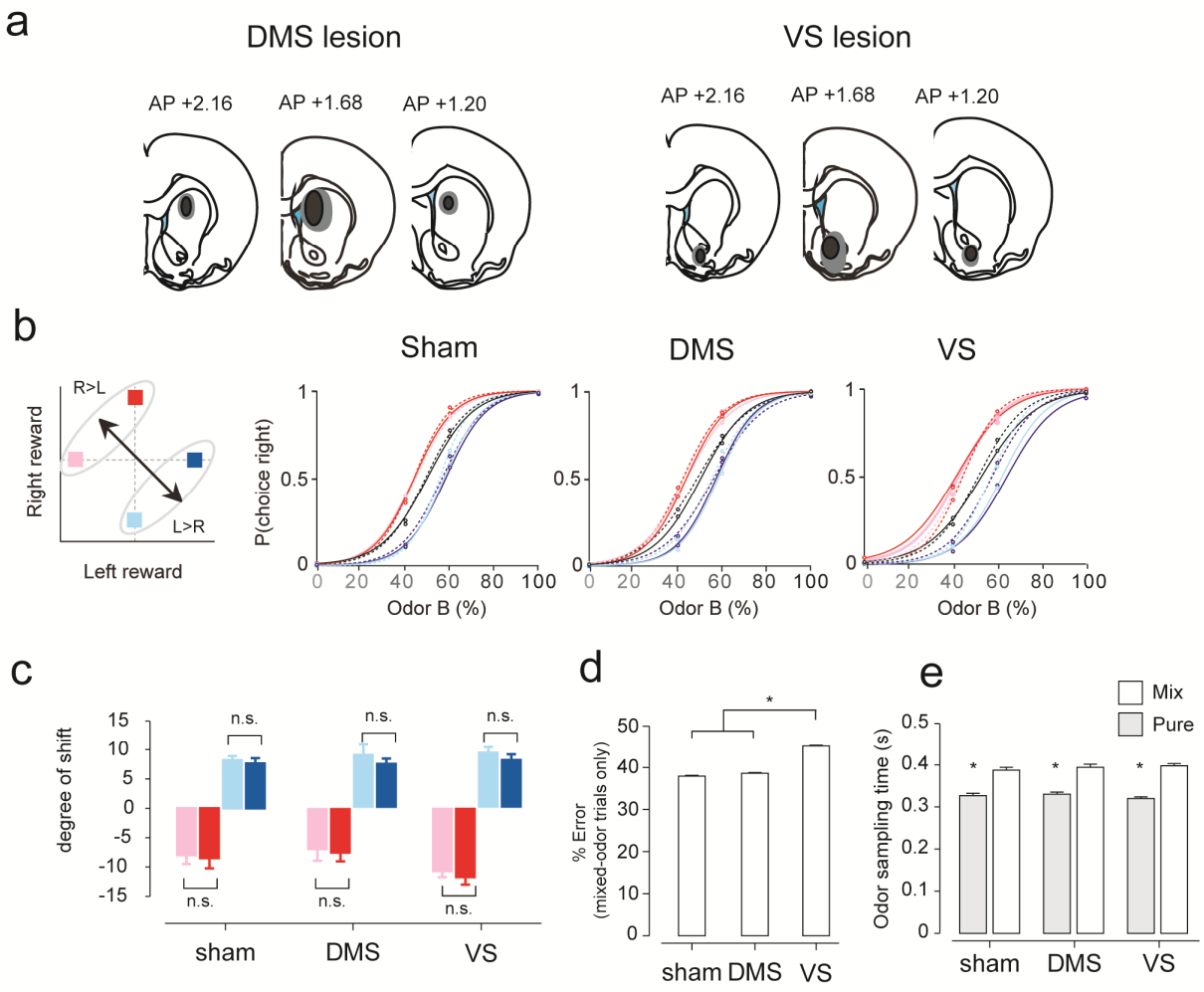

Figure 1.6 Influence of selective lesions in the DMS and VS on choice behavior.
(a) Lesion sites. Black: minimal extent of lesions. Grey: outer extent of lesions.
(b) Psychometric performance curves of different lesion conditions. Dotted lines indicate pre-lesion performance; solid lines indicate post-lesion performance. **(c)** Shifts in psychometric curves across lesions conditions. **(d)** Percent errors for mixed-odor trials. **(e)** Odor sampling time of animals in different conditions are similar across conditions.

DMS, but not VS, is critical for net-value-dependent energizing of actions

Next, we examined the effect of lesions on trial initiation time. We first plotted the trial initiation time across blocks (mean of each block's overall median trial initiation time) for each lesion condition and noticed no significant difference across conditions. Thus, all animals initially appeared to appropriately regulate their vigor across blocks (**Fig. 1.7a**). Next, we analyzed trial initiation time of individual rats after different trial types using the vector representation method described above. **Figure 1.7b** shows that trial initiation vectors for all trials and those divided into correct and error were within 22.5 degrees of 225^o in all example rats, suggesting that initiation time was net-value dependent (**Fig. 1.7b**, individual examples from each lesion condition). However, vectors obtained separately for left and right reward trials revealed striking differences across lesion conditions (**Fig. 1.7b,** after left and after right trials). In the sham animals, trial initiation vectors for both left and right trials were within 22.5^o of 225^o , indicating net value-dependence (**Fig. 1.7b** sham: left trials θ_L= 216^o [lower bound: 191^o, upper bound: 243^o]; right trials θ_R= 225^o [lower bound:192^o, upper bound: 265^o]; **Fig. 1.8a population**: left trialsθ_L= 219^o±5.8 , right trials θ_R = 236^o±7, mean±s.e.m).

For the DMS-lesioned rats, however, neither of the trial initiation vectors for left or right trials showed net value dependence. Vectors for trial initiation following left trials depended on left value alone (**Fig. 1.7b** DMS example: θ_L = 190^o [lower bound: 175^o, upper bound: 205^o]) while trial initiation after right trials depended on right value alone (**Fig. 1.7b**, DMS example: θ_R = 289^o [lower bound: 268^o, upper bound: 314^o]).

Figure 1.7 DMS lesions impaired net value-dependent energizing of trial initiation

(a) Trial initiation time across blocks across different lesion conditions.

(b) Vector representation of trial initiation time across blocks, separated by previous trial outcome (correct, error, left reward, right reward) in example rats. Grey: 95% confidence interval. Dotted rings represent a scale of vector length (r =1).

(c) Vectors of trial initiation time for all animals separated by previous left and right reward trials.

(d) Difference in angle between left and right reward vectors across sham, DMS and VS lesioned populations.

(f) Effects of multiple previous trial outcomes on trial initiation time. Time constants are derived from exponential curves fitted to coefficients of multiple regression.

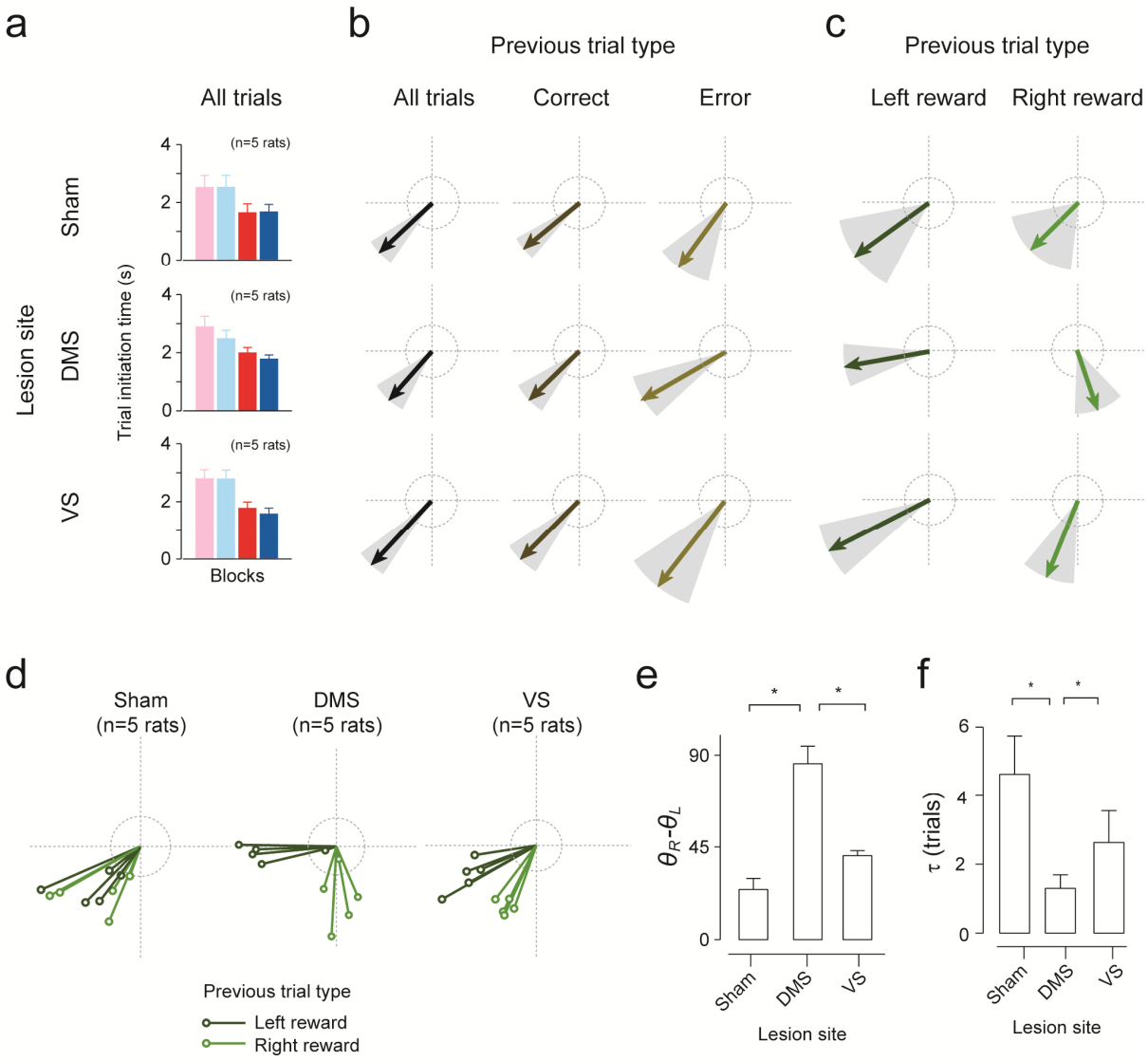

Figure 1.7 (continued)

The angle difference between left and right trial vectors in DMS-lesioned animals were nearly orthogonal to one another (**Fig. 1.7b** DMS example: $\theta_{R\text{-}L}$=99^o; **Fig. 1.7d:** $\boldsymbol{\theta_{R\text{-}L}}$=86^o±8.5; **Fig. 1.8a population:** left trials θ_L= 189^o ± 4.5, right trials θ_R= 275^o ±6.8, mean±s.e.m.) unlike sham-lesioned animals (**Fig. 1.7d:** $\theta_{R\text{-}L}$ = 24.2^o±5.35), indicating that their motivation does not actually depend on the net value of the block. Rather, trial initiation depended on the absolute value of the previous trial.

A less dramatic effect is seen in the VS-lesioned condition, where vectors were generally directed into the lower left-hand quadrant, and few vectors' confidence intervals crossed the 90^o or 180^o boundary (**Fig. 1.7b** VS example: left trials θ_L= 207^o [lower bound: 193^o, upper bound: 222^o], right trials θ_R =248^o [lower bound: 230^o, upper bound: 268^o]; **Fig. 1.8a population:** left trials θ=203^o±3.7, right trials θ=243^o±2.9, mean±s.e.m.). Vectors for left and right trials were more widely separated in VS animals (**Fig. 1.7b** VS example: $\theta_{R\text{-}L}$ = 41^o; **Fig. 1.7d** $\theta_{R\text{-}L}$: 40.7^o±2.4) than for sham animals, yet not orthogonal as in the DMS condition. This indicated that previous reward size had a stronger influence on VS-lesioned animals' trial initiation than sham animals, but not nearly to the extent as in DMS-lesioned animals.

Lastly, we examined the effects of multiple previous trial outcomes on trial initiation time via multiple regression as we had done above and fitted an exponential for the regression coefficients for previous trial outcomes. We found that the influence of the immediately previous trial on initiation time was significantly shorter for DMS-lesioned animals (**Fig. 1.7f,** sham: 4.2±1.1 trials; DMS: 1.2±0.4; VS: 2.9±1.2, mean ± s.e.m.). In total,

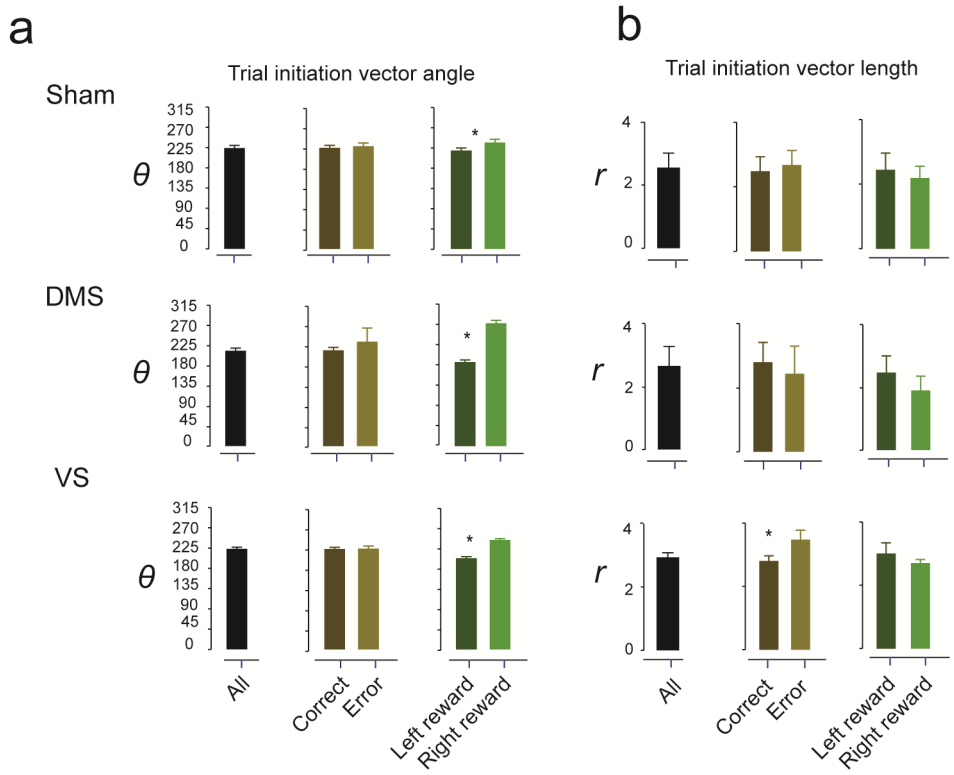

Figure 1.8. Population data: DMS lesions impaired net value-dependent modulation of trial initiation.
Trial initiation times are represented as vectors with polar angles and amplitudes projected onto value space. **(a)** Polar angles (mean ±s.e.m.) **(b)** Amplitudes (mean±s.e.m.).

these results demonstrate that the DMS has a larger role than the VS for integrating the total value of one's options, which is necessary for promoting state dependent modulation of general motivation.

Finally, we display the boxplots of trial initiation times for single animal examples as well (**Fig. 1.9**). For example, observe the trial initiation latencies following right trials in the DMS-lesioned rat. If trial initiation latencies depended on net value alone, trial initiation latencies would be long in both pink and cyan blocks and short in both red and blue blocks (as in the sham case, **Fig. 1.9**). However, we observe that the following right reward trials, the DMS-lesioned rat displays shorter latencies in cyan than in pink blocks. This is because in cyan blocks, the absolute size of right reward is medium sized. Similarly, in blue blocks, where the absolute size of right reward is medium, we observe longer latencies than in red blocks, despite both being of high net value.

DMS-lesioned animals' vigor to perform was immediately energized by large outcomes, were slow after small outcomes, and intermediate after medium sized rewards. In total, the DMS lesions did not seem to impair tracking of absolute values or relative values. Further, these results demonstrate that the DMS has a larger role than the VS in integrating the net value of options, which is necessary for promoting flexible, net-value dependent modulation of response vigor.

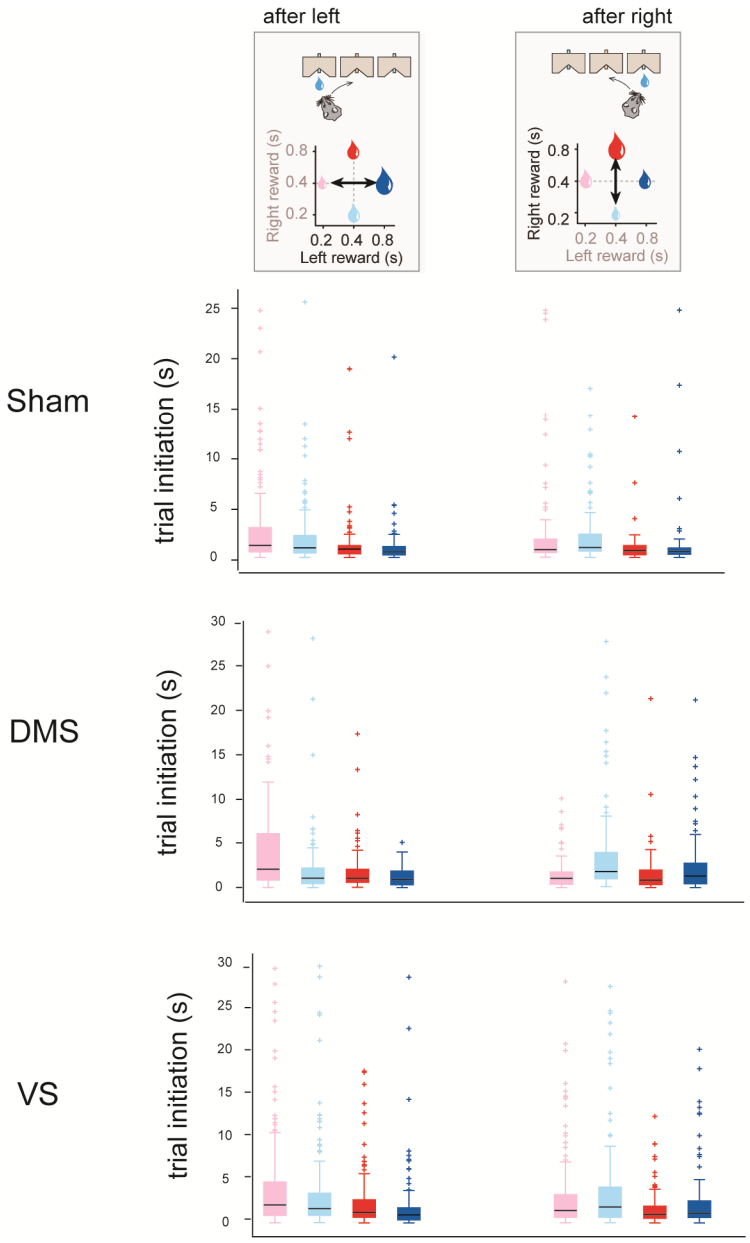

Figure 1.9 Boxplots of example sham, DMS, and VS lesioned trial initiation time across blocks, separated by previous left and right reward trials.

Discussion

Discussion of behavioral results

Our self-paced, value-based decision making paradigm allowed us to dissociate and independently measure both the energizing and directing effects of reward value on behavior in a unified task. An advantage of our self-paced task over tasks that tightly controls the timing of trial epochs is that we not only examine choice behavior but also the degree of the animal's engagement with the task. Further, we dissociated these two orthogonal aspects of decision making simultaneously in one task.

By simply manipulating the reward values of the animal's options, rats adapted the strategy of modulating their overall drive according to the net value of the block, independent of relative value. Oppositely, they chose actions based on their relative value, independent of net value. Further, rats initiated trials faster in states of high net value regardless of previous trial outcome. Whether they received large, medium, small, or no reward at all, they performed generally faster in high net value blocks, in order to exploit the task in states of high average reward rate.

Past experimental and computational studies on motivation have indicated that average or net reward rate drives motivation (Killeen, Hanson et al. 1978; McClure, Daw et al. 2003; Niv, Joel et al. 2006). A key consequence of Niv's model of motivational drive is that the effect of net reward rate on response rates should be outcome-general, rather than outcome-specific. The higher the net reward rate, the higher the cost of acting slowly for *any*

action because the cost of sloth delays *all* future rewards. Thus, when the opportunity cost of time is high, animals should exploit the task by performing more trials and faster, regardless of proximal outcomes. Our results provide experimental support that net value can globally energize task performance.

Our results are reminiscent of the Hullian generalized drive in that we may describe a generalized effect of net value on motivation. However, a major criticism of Hull's drive theory is that the internal 'drive' may simply be due to 'ignorance of the stimulus.' Here, we may be more cognizant of the stimulus that causes the outcome-general effects of response vigor, which is the net value of the animal's options. Further, we show that 'generalized drive' can be driven by external reinforcement, whereas Hull envisioned the buildup of a singular internal drive caused by physiological deprivation as the means to energize actions.

Discussion of Lesions

Lesions of the DMS and VS affected choice biases and vigor of rats in different ways. Perhaps surprisingly, the VS had little effect on the energizing aspect, whereas the DMS seemed to be most critical. In the past, the VS has had a 'privileged' position in motivational theory as a putative site of integrating motivation to action (Wise, 2004). This is largely due to the VS' famous role in motivating addictive behaviors while the DS has been more closely linked to the motor-related deficits of Parkinson's disease and to the formation of habits (Everitt and Robbins 2005). However, there is evidence that the DMS is most critical for allowing flexible, goal-directed behavior (Yin, Ostlund et al. 2005;

Ragozzino 2007). In our particular task, block changes happen every 40-60 trials, requiring animals to quickly adapt their motivational strategies to maximize reward and minimize energetic costs of performance. In this case, the DMS may be better suited to track such dynamic changes in net value. This is supported by physiology results demonstrating that DMS but not VS (Kimchi and Laubach 2009) nor DLS (Thorn, Atallah et al. 2010) neurons flexibly adapt to changes in stimulus-value or action-outcome contingencies. While our results support the idea that the DMS is required for flexible, goal-oriented behavior, we show that DMS lesions more profoundly affect flexible changes in motivation strategies than choice biases. This suggests a larger role for the DMS in the selection of motivational drive than previously thought.

Lesions of the DMS interestingly caused animal's motivation to become dependent on proximal outcomes. Rather than speeding their trial initiation for all trial types during blocks of high net value, the animals responded to the absolute value of the immediately preceding trial's outcome. This suggests that the DMS may normally be critical for integrating the values of the animal's options, or computing net value.

Another interpretation of our lesion results in the DMS is that striosomes or patches in the striatum may contribute to the selection of motivation. It's known that the head of the caudate, or anterior, dorsomedial parts of striatum is relatively more enriched in striosomes than other regions (Desban, Kemel et al. 1993). Thus, it may be worthwhile to further distinguish the role of these structures in the regulation of motivation.

The lesions performed above were done bilaterally, rendering animals to simply respond to the immediately preceding reward value after both left and right reward. Thus, it is possible that unilateral lesions could have yielded unilateral effects, for example, rendering initiation time after contralateral choices only (for example) to depend on the absolute value of reward. However, past studies have shown that while unilateral DLS lesions cause strong spatial bias contralateral to the side of lesion, DMS lesions cause much smaller degree of spatial bias (Brown and Robbins 1989).

Finally, it's well known that the DMS heavily interacts with the medial prefrontal cortex, together forming 'associative' cortico-basal ganglia loops. Thus, the DMS may not act in isolation to regulate motivation. It remains to be seen how other components of the associative cortico-basal ganglia loops such as mPFC participate in drive regulation and how their roles differ from that of the DMS. Moreover, it remains to be seen what role dopamine plays in the DMS's function in energizing actions. Past studies have shown that dopamine depletion in DMS but not VS impairs initiation of goal-directed responses (Amalric and Koob, 1987). Further, DMS also reciprocally projects to the SNc, suggesting lesioning the DMS may significantly affect the activity of dopamine neurons of the SNc. Future studies examining DMS' effects on dopamine phasic or tonic activity may provide insight on the function of striato-nigral loops.

Methods

All procedures involving animals were carried out in accordance with NIH standards and approved by the Harvard University Institutional Animal Care and Use Committee (IACUC). All values were represented by the mean ± standard error (SE) unless otherwise noted.

Behavior

15 male Long-Evans hooded rats were trained to perform an odor discrimination task for water reward (Uchida and Mainen 2003). Behavioural testing was controlled by custom software written in Matlab (Mathworks) using data acquisition hardware (National Instruments) to record the port signals and control the valves of the olfactometer and water-delivery.

Rats self-initiated each experimental trial by introducing their snout into a central port, which triggered odor delivery. Valid odor pokes were restricted to trials where animals delivered a single nose poke into the odor port and stayed in the odor port for at least 20 ms. After a variable delay, drawn from a uniform random distribution of 0.3–0.5 s, a binary mixture of two pure odorants, caproic acid and 1-hexanol, was delivered at one of 4 concentration ratios (100/0, 60/40, 40/6, 0/100) in pseudorandom order within a session. After a variable odor sampling time, rats responded by withdrawing from the central port, which terminated the delivery of odor, and moved to the left or right water port.Choices were

rewarded according to the dominant component of the mixture, that is, at the left port for mixtures A/B < 50/50 and at the right port for A/B > 50/50. For correct choices, reward was delivered by opening a water valve for the duration indicated by the block condition. 'Trial End'' is 4 seconds after closure of the water valve. The animal's trial initiation time is defined as the latency it takes for animals to poke back into the odor port after this 'trial end' time. Blocks were randomly interleaved within a session, which contained 5 to 9 blocks.

Lesions

After 6 weeks of training on the reward-manipulation task, rats were lesioned with ibotenic acid (250nl, 10mg/ml) in either the dorsomedial (AP 1.68, ML 2.0 and DV 4.5) or VS (AP 1.68, ML 1.5 and DV 7.4). After one week of recovery, animals were water deprived. Behavioral performance for the first 7-8 days after recovery was used for the behavioral analysis. The first and last 30 trials per session were excluded from the analysis for examining trial initiation, in order to examine steady-state behavior that is independent of satiety. The first 10 trials after each block transition was eliminated to exclude the effects of learning. Animals were perfused and stained with nissl as described before (Cury and Uchida 2010).

Analysis

We developed a method to represent how behavior activity was modulated by

changes across different reward blocks using a single vector in a polar coordinate system.

To obtain a behavioral vector, we regressed each animal's trial initiation time by the reward amounts of the left and right water ports, which varied across blocks. The values of the left and right choices, $Q_L$and Q_R, were defined by reward amounts (water valve duration; we confirmed the linear relationship between the valve durations and the delivered reward amounts).

We used F-test to examine whether behavior significantly modulated by blocks (1% significance). The F-test tests whether a proposed regression model as a whole fit the data significantly better than a simpler model (*Trial initiation* = β_0).

The amplitude (r) of the behavior vector was calculated by taking the square root of the sum of the square of the coefficients:

$$r = \sqrt{(\beta_L)^2 + (\beta_R)^2}$$

the polar angle (θ) was calculated by taking the four-quadrant arc-tangent of the coefficients:

$$\theta = \tan^{-1}(\frac{\beta_R}{\beta_L})$$

To determine how trial initiation time was modulated by reward size, we divided the polarplot into 8 segments of 45°: θ values of the behavior vectors falling between -22.5° and +22.5° represents a positive correlation between trial initiation and size of left reward;

behavior vectors whose θ fall between 22.5^{o} and 67.5^{o} represents a positive correlation between trial initiation and net-value, θ between 67.5^{o} and 112.5^{o} represents a positive-correlation between trial initiation and right-reward, θ between 112.5^{o} and 157.5^{o} represents positive correlation between trial initiation and right>left, θ between 157.5^{o} and 202.5^{o} represents negative correlation between trial initiation and left reward, etc.

CHAPTER 2

Value coding in the dorsomedial and ventral striatum

Introduction

Given stock options A and B, which one do you prefer? The field of decision making and neuroeconomics have made significant progress on this problem of 'action selection', or how one chooses a winner among a set of alternative choices. The process of action selection is theoretically separated into at least two main steps: valuation (representation of the value of each potential action on a common scale, or absolute valuation) and action selection (comparing the values of these choices relative to one another, or relative valuation) (Rangel, Camerer et al. 2008; Kable and Glimcher 2009; Lee, Seo et al. 2012). In support of this model, recent studies suggest that certain brain regions may participate in specific steps of the process: while the dorsomedial striatum (DMS or caudate) may participate in absolute valuation, cortical regions such as lateral intraparietal cortex (LIP) computes relative-values for selection (Platt and Glimcher 1999; Samejima, Ueda et al. 2005; Kim, Hwang et al. 2008; Lau and Glimcher 2008; Kim, Sul et al. 2009; Kimchi and Laubach 2009; Wunderlich, Rangel et al. 2009; Rorie, Gao et al. 2010; Cai, Kim et al. 2011).

Decision making, however, goes beyond selecting the better among alternatives. During an economic downturn, for example, one may decide to refrain from investing in *any* stock, despite the fact that A is better than B, and during times of prosperity, one may invest more frequently, perhaps ending up acquiring both. This is because one's motivation to exploit or invest in the stock market fluctuates according to changes in net expected return.

Several brain regions have been implicated in encoding the sum of action values or average reward value. Many of these regions overlap with those thought to be involved in valuation or selection such as the ACC (Seo and Lee 2007; Kolling, Behrens et al. 2012), LIP (Seo, Barraclough et al. 2009), VS, (Ito and Doya 2009; Cai, Kim et al. 2011), DS (Cai, Kim et al. 2011) and ventral pallidum (Ito and Doya 2009). These net value-related signals may be used to determine whether to engage in a task or how vigorously to perform, as well as be used to guide foraging behavior (Hayden, Pearson et al. 2011). In total, many cortical and sub-cortical brain areas participate in the representation of value-related information, but few have comprehensively examined how absolute, relative, and net value are encoded. Moreover, it is still unclear how different components of the cortico-basal ganglia loops differ from each other with regard to value representation.

Here, we comprehensively examine absolute, relative, and net value coding in the DMS and VS rats.

Results

We trained four rats on the self-paced, two-odor discrimination task described in Chapter 1, where a binary mixture of odorants signaled to the animal to obtain a water reward at the left or right reward port. We manipulated reward amounts of left and right water ports systematically in blocks of 40-60 trials (**Fig. 2.1)** in two task types. Two animals performed the task shown in **Figure 2.1a**, top panel (Samejima et al., 2005) and two performed the task shown in **Figure 2.1a**, bottom panel (Rorie et al., 2010). After rats were trained on the task, we implanted into the left hemisphere of each rat a custom-made microdrive consisting of 12 tetrodes each. We recorded 522 neurons, with 364 from DMS and 158 from VS in rats as they performed the task (n=4 rats; 103.5± 22.7 and 39.5 ± 56.9 neurons in DMS and VS per animal; mean ± S.D.).

By simply manipulating the reward amounts of the left and right choices systematically in blocks of trials, we showed that relative values biased the animals' choice behavior while net values modulated trial initiation time. This pattern held true for both task types.

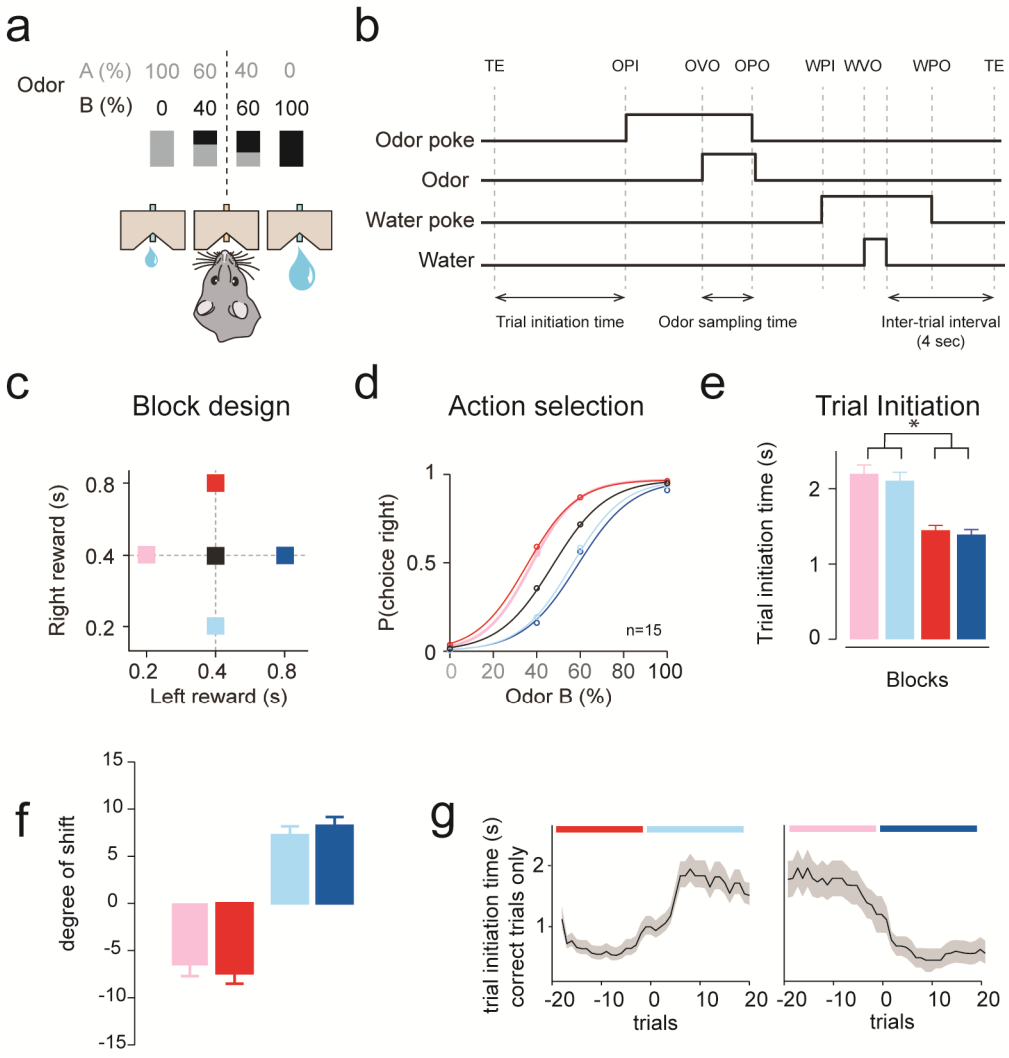

Figure 2.1 Behavioral performance during neuronal recording of two different task types: relative value bias choices, net value drives trial initiation time
(a) Block-wise manipulation of left and right reward value, two block types (top and bottom)
(b) Psychometric performance curves in different blocks (colors correspond to blocks in **a**
(c) Trial initiation time per block (mean ± s.e.m.)

Animals performed with near perfect accuracy for easy, pure odor trials, indicating that they wait until they have sampled the odor to make their choice selection. Although animals do not finalize their choice of which water port to go until after odor onset, the block-wise structure of the task allows the animal to know the action value of left or right choice by the beginning of each trial (minus the first few trials of each block). (Note: although we use the terms 'action value', we do not dissociate it from 'outcome value' in this study). Thus, even before the trial has initiated or before odor cue is delivered, the animal knows the expected value of the block. Therefore, neural activity in the earliest epochs of each trial can be used to bias action selection.

Moreover, we examined each neuron's time of maximum firing and found that as a population, the inter-trial epoch was a time of high neural activity, particularly around 'odor-poke-in', or when the animal pokes into the odor port to initiate the trial (30%, or 152/522 cells had maximal firing during this inter-trial interval)(**Fig. 2.2**). These results, in conjunction with our lesion results described in Chapter 1 that implicated the DMS in modulation of trial initiation, led us to first examine neural activity related to expected value during the pre-trial initiation period. Additionally, value-related activity during this early trial epoch may be used to later guide action selection. In the following section, we focused our neural analysis on the epoch 0-300 ms before 'odor-poke-in' (when the animal self-initiates by poking into the odor port).

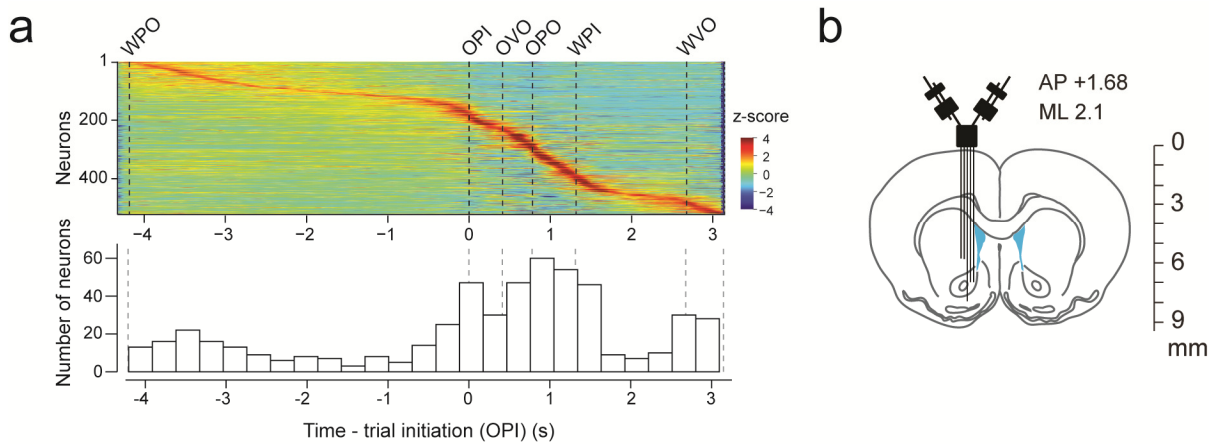

Figure 2.2 Tetrode recording of dorsomedial and ventral striatum
(a) Peri-event time histograms (PETH) of all recorded neurons sorted by time of peak firing (z-scored). Due to trial-by-trial variability of the timing of task epochs, PETHs are time-warped and aligned by the average time of task epochs. **(b)** Recording configuration. Abbreviations: WPO: water poke out; OPI: odor poke in; OVO: odor valve on; OPO: odor poke out; WPI: water poke in; WVO: water valve on.

To quantify how the neural responses during the pre-initiation epoch are modulated by block-wise changes in value, we projected each neuron as a vector onto "value space," by regressing the firing rate of each neuron with left and right value (Q_L and Q_R, respectively).

$$\textit{Firing rate (spikes/s)} = b_0 + b_L \cdot Q_L + b_R \cdot Q_R \qquad (2)$$

The polar angle of each neuron's response allows us to map them to their respective decision making processes: neurons positively or negatively correlated with left value have θ of 0^o or 180^o , respectively (**Fig. 2.3)**. Neurons positively or negatively correlated with right value are 90^o or 270^o, respectively. Such neurons participate in valuation, as they encode the 'action value' of either choice, but not both. In other words, these neurons track the expected value of one choice, irrespective of the value of the other choice, thus displaying 'menu-invariance'. Next, neurons encoding relative-values, Q_R-Q_L, are modulated towards 135^o or 315^o. A vector of 135^o indicates preferred firing in blocks when Q_R is relatively greater than Q_L and 315^o indicates the preferred firing when right is relatively greater than left. Orthogonally, neurons positively or negative modulated by net-values Q_L+Q_R, are 45^o or 225^o, respectively. Value-coding neurons are defined as those that showed significant modulation in any direction in this value space (P<0.01, F-test).

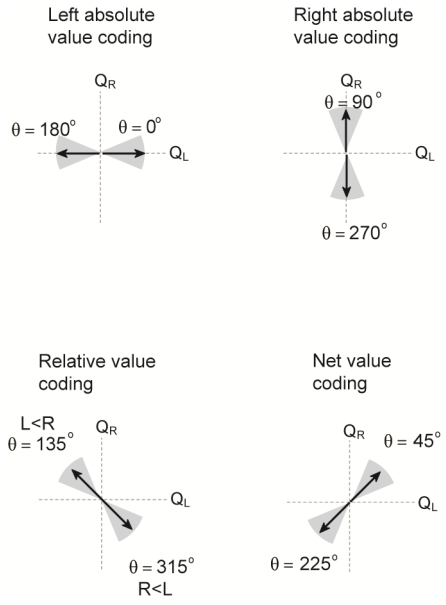

Figure 2.3 Classification of neural responses based on the polar angle of firing rate vectors

Figure 2.4a-c shows example peri-event time histograms of different types of value coding neurons and **Figure 2.4d-f**, their corresponding vectors that show the direction of value modulation. For example, **Figure 2.4a** shows a neuron whose firing rate is highest in pink blocks, lowest in blue blocks, and intermediate in red and cyan blocks. In other words, the firing rate is negatively correlated with changes in left action value, but invariant to changes in right action value. This neuron is represented as a vector that is modulated across the horizontal axis and has $\theta = 194^{o}$ (**Fig. 2.4d**), indicating that it is a negatively modulated by the absolute-value of the left option. In contrast, **Figure 2.4b** shows a neuron that is modulated by relative-value, as it prefers to fire in blocks where left reward is relatively bigger than the right and has a vector of $\theta = 331.5^{o}$ (**Fig. 2.4e**). A net-value dependent neuron is shown in **Figure 2.4c**, whose vector points in the lower left-hand quadrant with $\theta = 219^{o}$ (**Fig. 2.4f**).

The DMS predominantly encodes net value before trial initiation

We performed the vector analysis for all recorded neurons in the DMS and VS and then examined the distribution of absolute, relative, and net value coding neurons in either region. We plotted vectors for neurons that were significantly modulated by value across blocks during the pre-trial initiation epoch ($P<0.01$, F-test) separately for DMS and VS (**Fig. 2.5a**). In all, 31% (113/364) of DMS neurons were significant and 22.8% (36/158) were in the VS. Although the DMS contained a higher proportion than the VS, the difference was not significant ($P = 0.05$, χ^2 test).

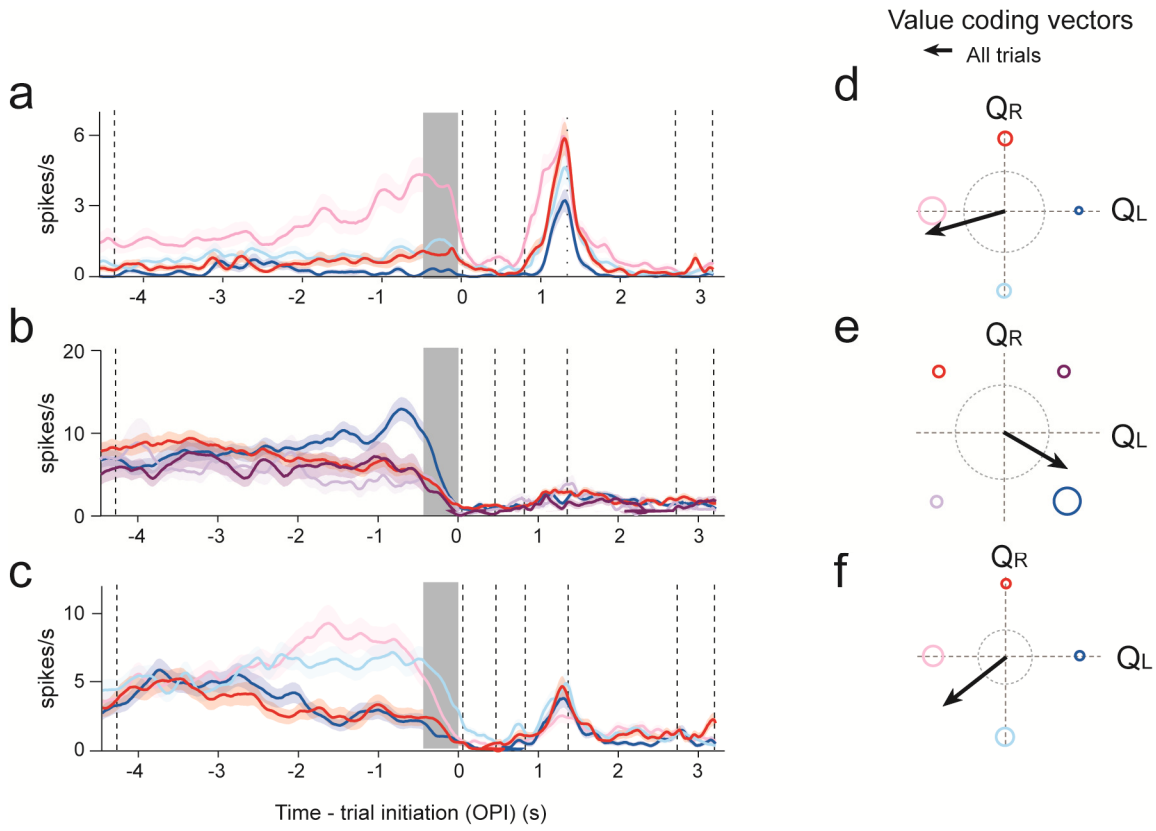

Figure 2.4 Example peri-event time histograms of value-coding neurons
(a-c) Peri-event time histograms of example neurons. Colors refer to the blocks in Fig. 2.1.
(d-f) Vector representation of neural activity during the pre-trial initiation epoch.

The distribution of the vectors around the polar-coordinate of the two striatal regions, however, appeared different. In the VS, the proportion of absolute and net value coding neurons were represented significantly above chance, but relative value coding neurons were not (**Fig. 2.5b, bottom;** $P > 0.0125$, Bonferroni correction for proportion of relative value coding neurons). Finding such low proportions of relative-value coding neurons in VS is consistent with a previous study in monkeys (Cai et al., 2011). Further, the distribution of left-absolute, right-absolute, relative, and net-value coding neuron types was not significantly different from uniform ($P = 0.57$, goodness of fit test against uniform).

On the other hand, all categories of neuron types (absolute, relative, and net value coding) were significantly represented above chance level in the DMS ($P < 0.0125$, binomial test)(**Fig. 2.5b**). Yet, the distribution of these categories was significantly different from uniform ($P = 0.0077$, χ^2 test goodness of fit against uniform). Strikingly, the non-uniformity was due to the predominance of a single category: neurons encoding net value (**Fig. 2.5b**). The proportion of net value coding neurons was the one category that significantly deviated from what is expected from a uniform distribution ($P = 0.0028$, binomial test). Therefore, the DMS represents all coding types, with net value coding neurons being most dominant.

One advantage of our method of classifying neurons as vectors is that in addition to quantifying the number of neurons per category, as classified by θ, we can also take into account the strength of representation, or the amplitudes, r. We computed resultant vectors for each category of neurons with their 95% confidence boundaries (bootstrap) and projected

them onto the polar coordinate (**Fig. 2.5c**). After taking into account the amplitudes of the vectors, which corresponds to the strength of neural modulation, the DMS still predominantly represents net value over other value coding types.

Finally, we examined whether the activity of net value coding neurons in the striatum were significantly affected by the immediately preceding trial's outcome. We split trials into left and right reward outcome trials and performed the vector analysis for the firing rate during the pre-initiation epoch (**Fig. 2.6a-b**). The majority of neurons (>80%) showed net-value dependent firing modulation regardless of previous trial's outcome of neurons. As a population, the angle between vectors for firing rate immediately after left reward and after right reward trials were not statistically significant (paired T-test, θ_L vs θ_R, $P>0.05$). Moreover, we regressed firing rate with net value and previous trial outcome and found that 94% (51/54) of the neurons were significant for net value while (7/54) 13% were significant for both net value and previous trial outcome while only 3.7% (2/54) were significant for previous trial outcome alone.

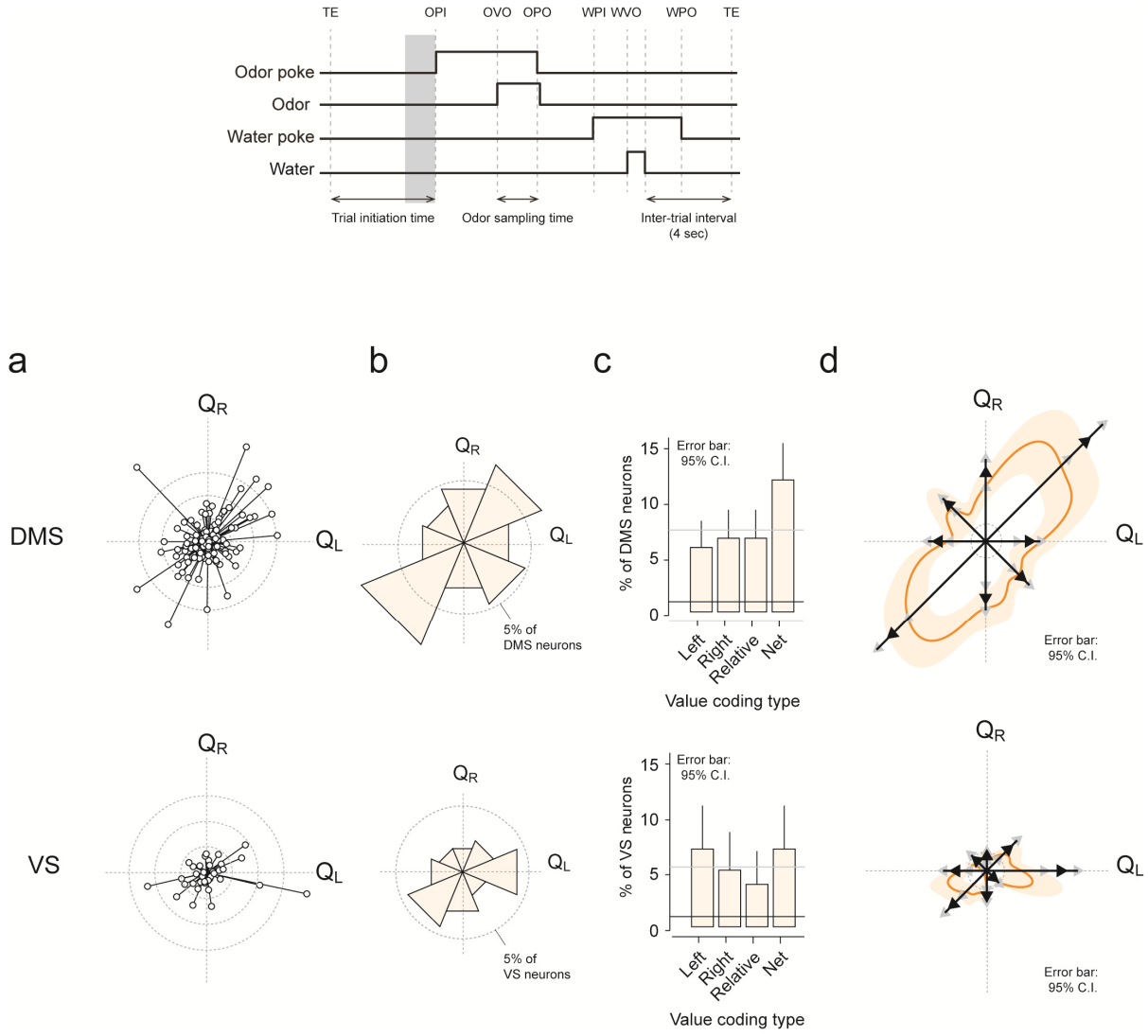

Figure 2.5 DMS encodes net value before trial initiation
(a-c) Top: DMS value coding neurons; Bottom: VS value coding neurons. **(a)** Polar representation of all value coding neurons (P<0.01, F-test). Inner dotted ring: r=9, middle ring: r=16, outer ring, r=24. Q_L= left value; Q_R= right value. **(b)** Circular histogram depicting the proportion of neurons in each category. Dotted ring, 5%. **(c)** Distribution of value coding types. Mean ± 95% confidence intervals. Upper dotted lines: expected uniform distribution of value coding neurons. Lower lines: expected fraction of chance identification. **(d)** Vectors: resultant vectors of each value coding type (black arrowheads). Grey arrowheads represent the boundaries of the 95% confidence interval. Orange: smoothed vector representation (thick orange line). Mean±s.e.m. (thick line and shading, respectively).

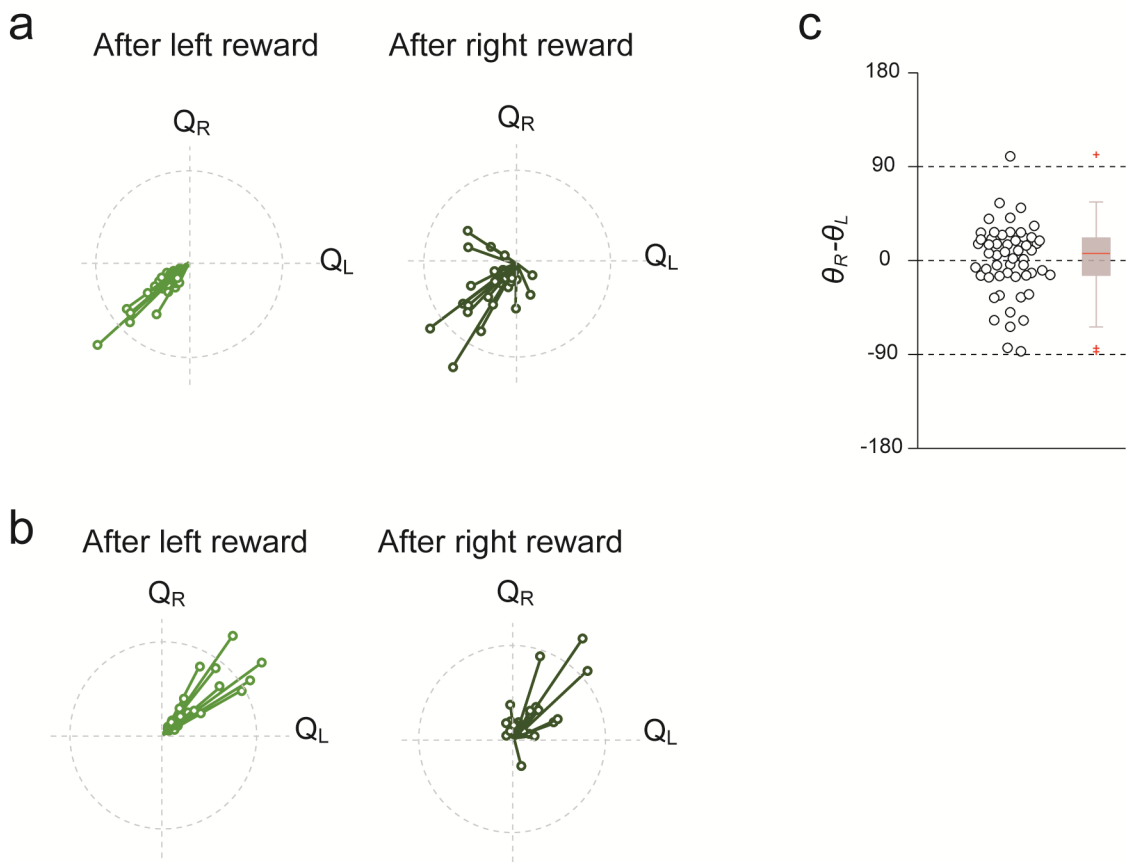

Figure 2.6 The outcome of immediately preceding trial does not significantly affect the activity of net value coding neurons
(a) Neurons having negative correlation with net value: vectors representing direction of value modulation are plotted separately for trials following left and right reward.
(b) Same as Fig. 8a for neurons having positive correlation with net value.
(c) Difference in angle between vectors for firing rate after left and right reward trials.

In summary, the DMS is enriched with net value-related activity just before trial initiation. Although it contains significant numbers of absolute and relative value neurons, net value representation dominates. Thus, the DMS may be more involved in energizing behavior than previously thought. The VS, on the other hand, has relatively fewer value coding neurons during the pre-initiation period. All but relative value coding neurons were represented in significant levels in the VS. Although the DMS contained significant numbers of relative-value coding neurons, relative value representation was still the weakest, leaving open the possibility that action selection-related activity may be more prominent downstream or elsewhere.

A popular model of decision making describes a hierarchical architecture, where absolute value representations in the DMS are read out and used by downstream areas to compute the relative-values needed for action selection (Daw and Doya 2006; Kable and Glimcher 2009). In support of this, past studies found that DMS neurons primarily encode the absolute 'action values' of future options prior to the manifestation of choice. However, our results show that net value representation in the DMS, which may be used to regulate motivation, is primary. Our results seem to contradict the current model of decision making because of two main reasons: 1) our experimental design comprehensively addresses and dissociates the representation of absolute, relative and net values, which have not previously been done and 2) our analysis methods differ from past studies in that they may be less biased towards classifying neural activity as absolute-value coding (see **Supplemental material**).

Value-related activity in DMS and VS after trial initiation

Value coding before odor stimulus

Next, we performed similar analyses during the delay period between odor-poke-in (trial initiation) and odor valve onset. Similar to the pre-odor-poke-in epoch, the animal had not yet made its choice of which water port to approach but it knows the expected values of its options (due to the block-wise nature of the task) (**Fig. 2.7**). During the pre-odor period, significantly more neurons in the DMS were modulated by value than the VS (32.1% of DMS or 117/364; 20.9% or 33/158 VS neurons; P=0.009, χ^2 test). The distribution in the VS was non-uniform, where the majority of neurons were negative-state value coding. In the DMS, the distribution was not significantly different from uniform, although state value coding neurons were in the majority, comprising over 30% (34/117) of the four major value coding categories. Next, we obtained resultant vectors for each of the 8 categories of value coding neurons and show that similar to the pre-initiation period, the DMS seems to predominantly be modulated by net value. During this period, the VS was primarily *negatively* modulated by net value (**Fig. 2.7**).

Thus, we showed that early in the trial, striatal neurons represent the expected absolute, relative, and net values of the animal's options. The DMS in general contains a greater proportion of value modulated neurons than the VS. Moreover, DMS predominantly represents net value over the early trial epochs. However, the relative dominance of net-value representation is weakened during the pre-odor epoch as compared with the pre-initiation epoch. On the other hand, the VS encodes values uniformly before trial initiation and later, evolves to encode more negative-net value just before odor onset.

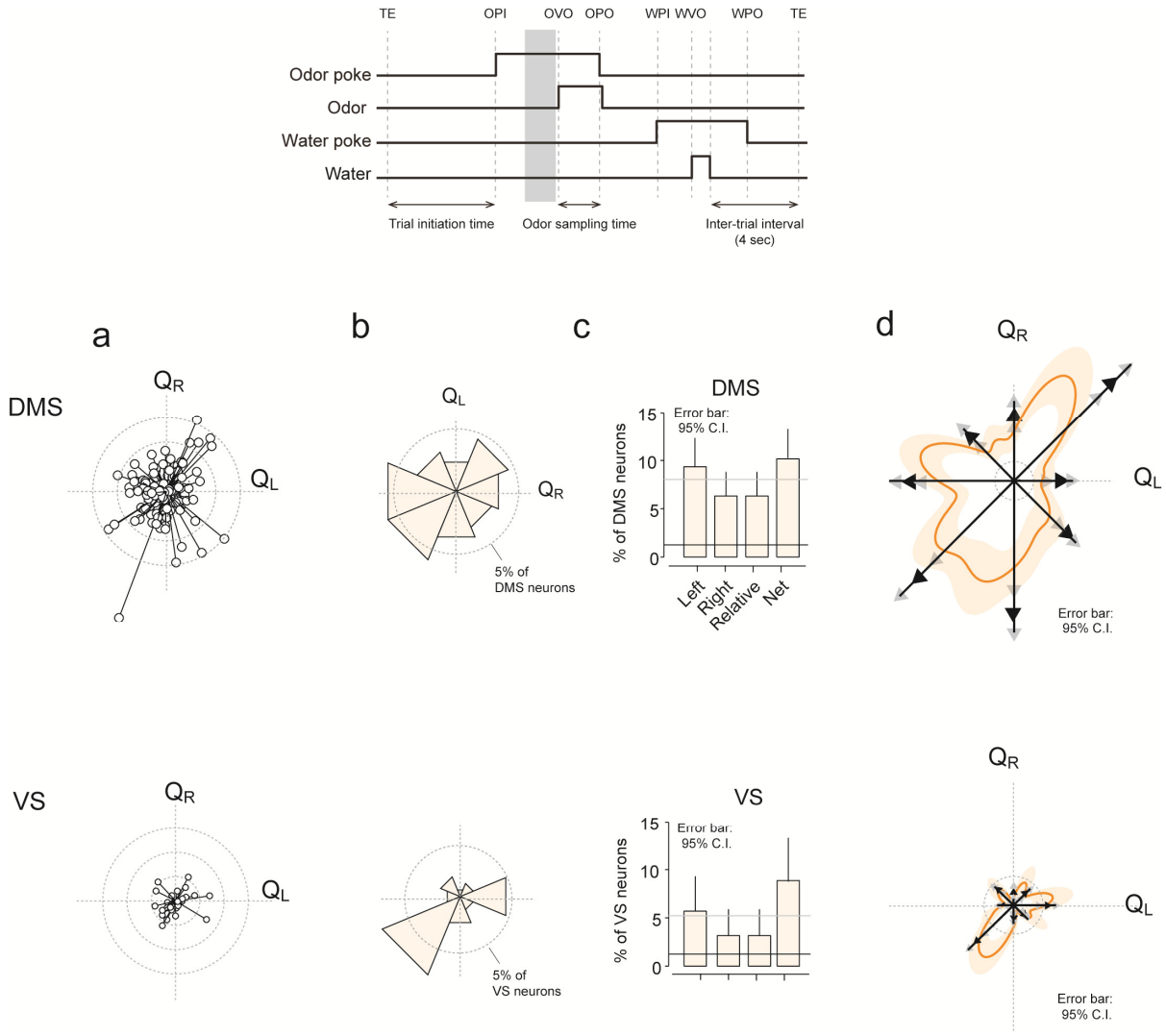

Figure 2.7 Value-related activity immediately before odor onset
(a-c) Top: DMS value coding neurons; Bottom: VS value coding neurons. **(a)** Polar representation of all value coding neurons (P<0.01, F-test). Inner dotted ring: r=9, middle ring: r=16, outer ring, r=24. Q_L= left value; Q_R= right value. **(b)** Circular histogram depicting the proportion of neurons in each category. Dotted ring, 5%. **(c)** Distribution of value coding types. Mean ± 95% confidence intervals. Upper dotted lines: expected uniform distribution of value coding neurons. Lower lines: expected fraction of chance identification. **(d)** Vectors: resultant vectors of each value coding type (black arrowheads). Grey arrowheads represent the boundaries of the 95% confidence interval. Orange: smoothed vector representation (thick orange line).
Mean±s.e.m. (thick line and shading, respectively).

Value modulation during odor stimulus

We next examined how neurons are modulated after the odor cue. We analyzed the time epoch 0-200ms during odor stimulus, before the overt expression of choice. Because of the arrival of various different types of odor stimuli during his time epoch, we first performed the analysis of variance (ANOVA) to compare the amount of neural activity related to odor stimulus, value, and their interaction in the DMS and VS. We found that there were significantly more neurons in the DMS than VS that encoded value, odor and value-odor interaction (DMS: 31.3% or 114/364 for value, or 13.7% 50/364 for odor, 10.4 or 38/364 for interaction; VS: 19.6% or 31/158 for value, 7.6% or 12/158 for odor, 5.1% or 8/158 for interaction; $P<0.05$ χ^2). Furthermore, significantly more neurons were modulated by value than by odor in both the DMS and VS ($P<0.01$,χ^2test). Therefore, it seems that during odor, activity modulated by value was more dominant than activity modulated by odor, and that value-related activity was stronger in the DMS than the VS. Our findings are compatible with Kim et al.'s findings that the DMS contains more signals related to value than the VS, and that these signals appear early in the trial (Kim et al., 2009).

To examine how neurons were modulated by value during the odor period, we separated trials into those where the animals subsequently chose left or right (which is also highly correlated with odor identity) and obtained vectors indicating their direction of value modulation in value space (**Fig. 2.8**).

For *left* (ipsilateral) choice trials in the DMS, *left* absolute value and *net* coding vectors were the only categories represented significantly above chance level. On the other

hand, for *right* (contralateral) choice trials in the DMS, *right* absolute value and *net* coding vectors were the only categories above chance level ($P < 0.0125$, binomial test). The distribution of value coding types in the DMS was significantly different from uniform for right (contralateral) choice trials but not for left choice trials (left trials: $P= 0.2982$; right trials: $P=0.0186$, χ^2 goodness of fit test against uniform)(**Fig. 2.8a-c**).

Similar to the DMS, the VS' distribution for right (contralateral) choice trials but not left choice trials was significantly different from uniform ($P= 0.9725$ left trials; $P=0.0107$ right trials, χ^2 goodness of fit test against uniform). For right choice trials, only net value and left absolute value coding neurons were represented significantly above chance ($P<0.0125$, binomial test) . However, none of the value coding categories were represented significantly above chance for left choice trials ($P>0.0125$, binomial test) (**Fig. 2.8 d-f**).

These results suggest that during stimulus but before animals revealed their choice, DMS neurons were modulated by the net value in addition to the absolute value of their subsequent choice, which is also called 'chosen value' (note that since choice and odor is tightly correlated, we cannot fully dissociate choice from stimulus). Very few relative value neurons were found during this period. In the VS, only right (contralateral) choice trials had significant proportions of value coding neurons, and the distribution was similar that during pre-initiation and odor epochs (negative net value and ipsilateral absolute value representation).

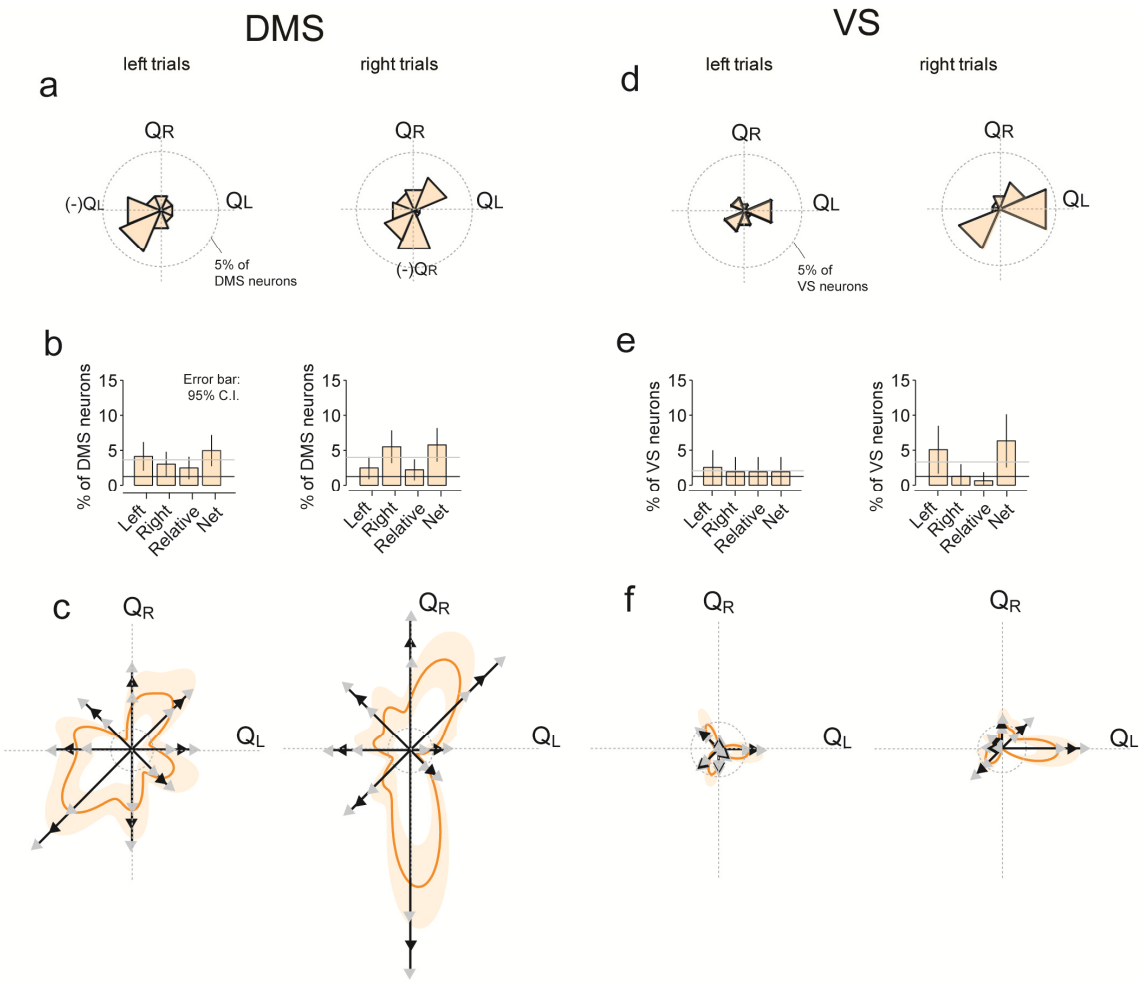

Figure 2.8 Net value and chosen value, but not relative value activity during stimulus presentation in DMS
(a-c) Statistically significant DMS neurons (left panels: left choice trials; right panels: right choice trials). **(d-f)** Statistically significant VS neurons (left panels: left choice trials; right panels: right choice trials). **(a,d)** Circular histogram of polar vectors. Dotted ring: 5%. **(b,e)** Distribution of value coding types. Upper lines: expected uniform distribution of value coding neurons. Lower lines: expected fraction of chance identification. **(c,f)** Vectors: resultant vectors of each value coding type (black arrowheads). Grey arrows represent the boundaries of the 95% confidence interval. Orange: smoothed population data.

Value modulation during choice: VS represents the relative value of chosen option

Next, we examined how neurons were modulated during choice (300ms after odor poke out) by obtaining value-modulated vectors separately for left and right choice trials (**Fig. 2.9**). For left choice trials in the DMS, all but relative value coding neurons were represented significantly above chance ($P<0.0125$, binomial test for all but relative value neurons). For right choice trials in the DMS, all neuron types are represented above chance level. Regardless of movement direction, the general trend in DMS is similar to that during odor onset: for *left* trials, many neurons were negatively modulated by *left* value (180^o) and positively by *net* value (45^o); for *right* trials, the majority of neurons were negatively modulated by *right* value (270^o) and positively by *net* value (45^o). Thus, neurons in the DMS encode either net value or the absolute value of the chosen action during movement.

In contrast, for left choice trials in the VS, relative value coding neurons were the *only* category significantly represented above chance. 46% of significantly modulated neurons for left choice trials preferred to fire when the left side was *relatively* larger than the left. Furthermore, 12 out of 13 relative value neurons were *left*-preferring for *left* choices. Thus, when animals make left (ipsilateral) choices, VS neurons fire when they are expecting the relatively better outcome (**Fig. 2.9d**). However, we did not observe similar relative-value modulation for right choices, where the distribution was uniform, although there were generally more neurons encoding absolute value of right choice. Furthermore, for right choices, all but net value coding neurons were significantly represented above chance level (binomial test). Thus, relative value signals in the VS dominate for ipsilateral choice.

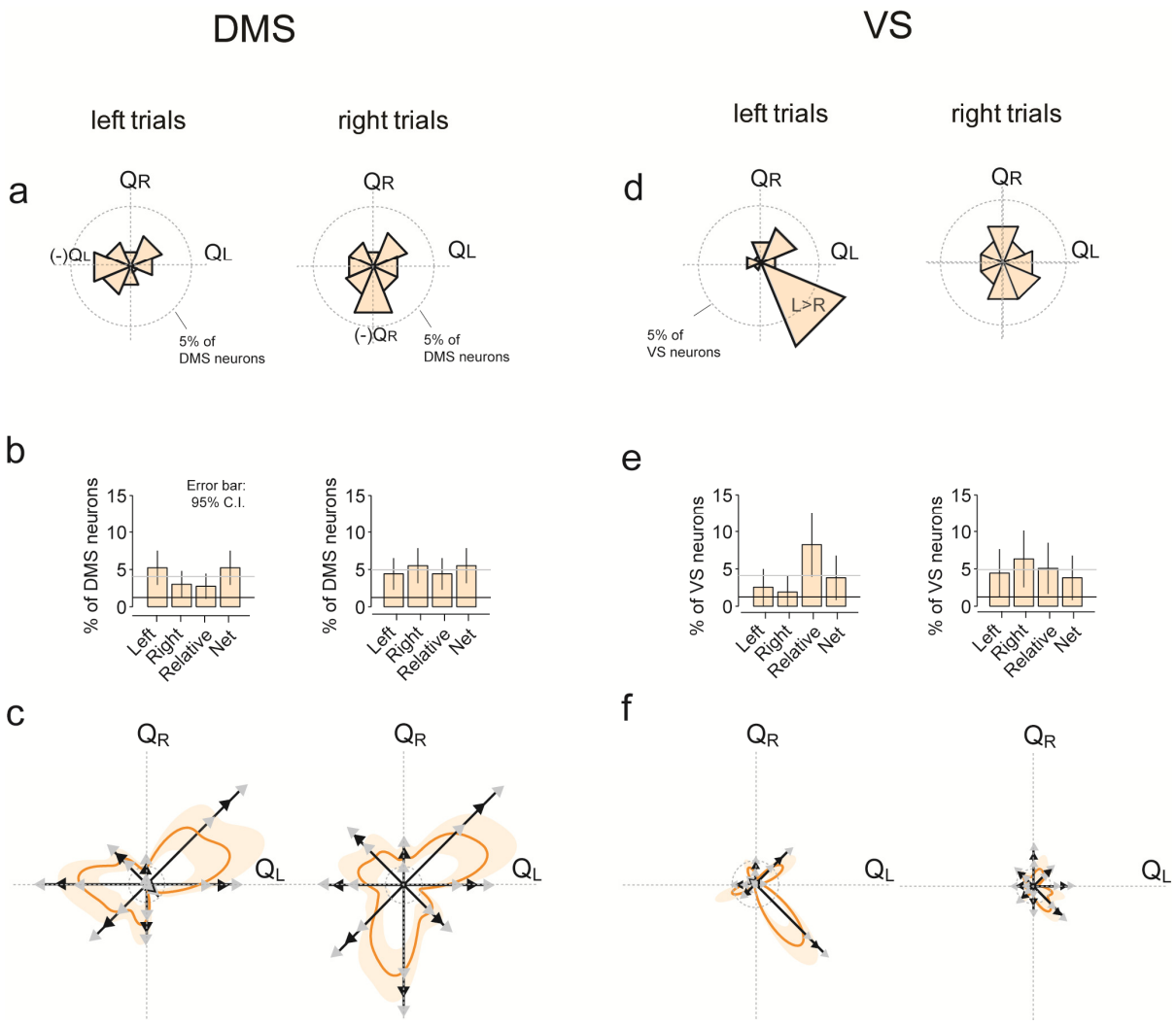

Figure 2.9 Value coding during choice: Relative-value coding predominates in the VS; Net and absolute value predominates in the DMS
(a-c) Statistically significant DMS neurons (left panels: left choice trials; right panels: right choice trials); **(d-f)** Statistically significant VS neurons (left panels: left choice trials; right panels: right choice trials). **(a,d)** Circular histogram of polar vectors. Dotted ring: 5%. **(b,e)** Distribution of value coding types. Upper lines: expected uniform distribution of value coding neurons. Lower lines: expected fraction of chance identification. **(c,f)** Vectors: resultant vectors of each value coding type (black arrowheads). Grey arrows represent the boundaries of the 95% confidence interval. Orange: smoothed population data.

We show an example neuron with left-preferring relative-value modulation for left choice trials (**Fig. 2.10**) as well as for right-choice trials, which had a smaller degree of left-preferring value modulation. In summary, from odor onset until movement, VS left-preferring relative value neurons formed the largest value coding subtype (compare circular histogram of left trials in **Fig. 2.9d** to all other circular histograms between odor and choice).

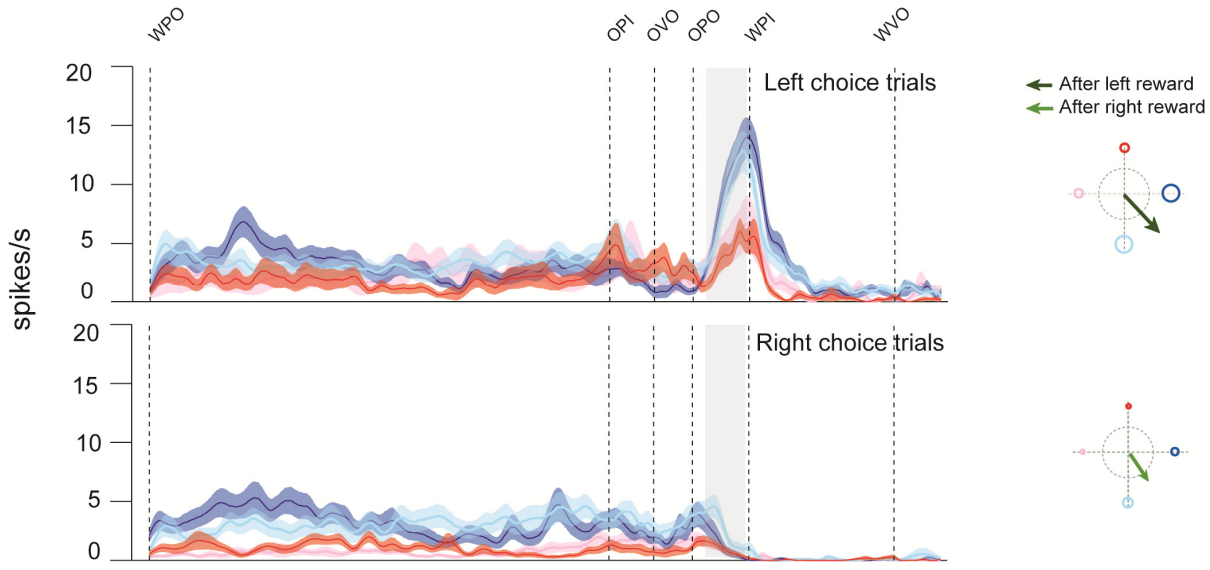

Figure 2.10
Peri-stimulus time histogram of a ventral striatal neuron that prefers to fire in left choice trials when the expected value of left reward is relatively bigger than right reward. Right panels: dotted circle: r=7.

Value modulation during reward consumption

Lastly, we examined value-modulated activity after reward consumption (using a 300 ms time window around water-valve-off for all correct trials) (**Fig. 2.11**). For left choice trials in the DMS, all but right absolute value (the un-chosen side's value) was represented above chance level (binomial test). In right trials, all but left absolute value (the un-chosen side) was represented above chance. In fact, the vast majority of significant neurons encoded the absolute value of the chosen side (**Fig. 2.11**: left choice trials encode left value; right choice trials encode right value).

The VS also predominantly represented the absolute value of the chosen side. For left trials, *only* left absolute value was represented significantly above chance level; for right trials, only right absolute value was represented significantly above chance level (**Fig. 2.11**) ($P<0.0125$, binomial test). Therefore in the VS, nothing but the chosen value (in absolute terms) was represented above chance level.

Interestingly, the DMS contained a larger proportion of *negatively* modulated vectors (46 positive vs. 74 negative, out of 364 neurons, $P=0.0052$, χ^2 test) while oppositely, the VS contained a larger proportion of *positively* modulated vectors (35 positive vs. 21 negative out of 158 neurons, $P=0.039$, χ^2test). In all, both VS and DMS encode chosen value during reward.

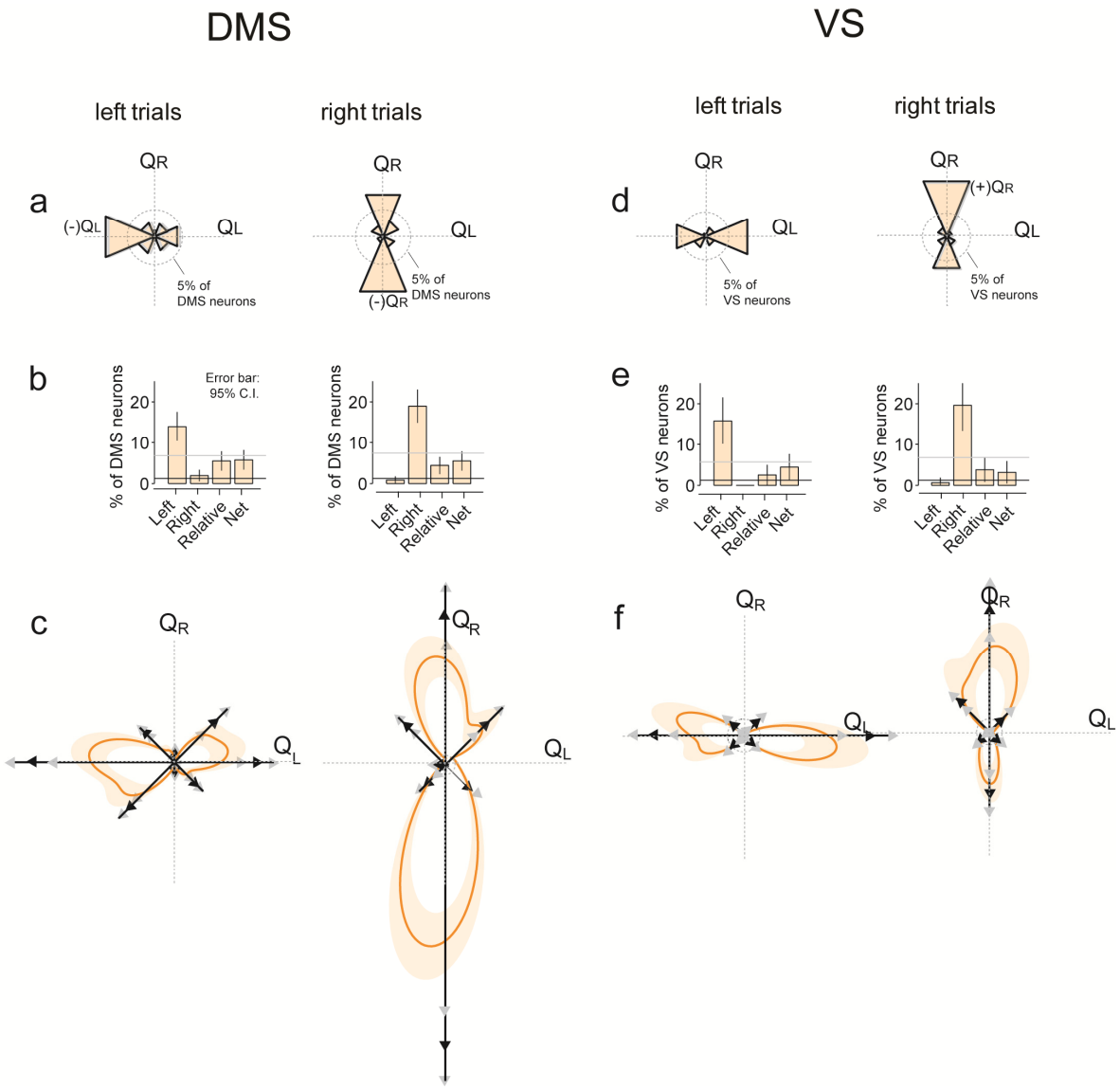

Figure 2.11 After reward consumption: VS and DMS encode chosen reward value
(a-c) Statistically significant DMS neurons (left panels: left choice trials; right panels: right choice trials). **(d-f)** Statistically significant VS neurons (left panels: left choice trials; right panels: right choice trials). **(a,d)** Circular histogram of polar vectors. Dotted ring: 5%. **(b,e)** Distribution of value coding types. Upper lines: expected uniform distribution of value coding neurons. Lower lines: expected fraction of chance identification. **(c,f)** Vectors: resultant vectors of each value coding type (black arrowheads). Grey arrows represent the boundaries of the 95% confidence interval. Orange: smoothed population data.

Discussion

In all, we show that the DMS predominantly represents net value in the early stages of every trial (before trial initiation and choice). Further, it generally contains more value modulated neurons during this early trial epoch than the VS. After action selection (odor onset until water port entry), the DMS maintains a large subset of net value coding neurons but additionally, neurons that encode chosen value (or stimulus) emerge. In the VS, the most striking difference from the DMS occurs during choice to the ipsilateral side: the majority of neurons prefer to fire when the chosen ipsilateral side is the relatively better option. Finally, after reward delivery, net value representation disappears, and the vast majority of VS and DMS neurons encode the absolute value of the actual reward. However, while VS is positively correlated with actual reward value, DMS contains more neurons *negatively* correlated with the value of obtained reward.

From prior to trial initiation until the animal reaches the reward port, the DMS is particularly enriched with net value representation. As seen in our behavioral results, animals 'exploit' the task during blocks of high net value by performing more trials and with faster latency, while in states of low net value, they are less engaged. These activities in the early epochs of the trial may reflect or be used to energize the motivation to perform.

Past studies indicated that the DMS primarily encodes the absolute action values of potential choices prior to the animal's manifestation of action (Samejima et al., 2005; Lau and Glimcher, 2008). Accordingly, the current model of action selection implicates that prior

to action selection, the DMS maintains the subjective value of every option on a common, menu-invariant scale so that downstream areas can compare options with one another. Our results show that prior to action selection, DMS is not restricted to absolute representations of action value. In fact, *net value* representation of options is most dominant. The VS, on the other hand, shows weaker representation of value in general. Moreover, its relative value representation is weakest, which is consistent with past studies in monkeys (Cai et al., 2011).

Our analysis further reveals how value representation is distributed rather continuously, as can be seen in (**Fig. 2.5a**), where vectors distribute all around the polar coordinate. Rather than simply forming distinct categories of value coding types, the population seems to encode diverse linear combinations of value. Future studies will be needed to understand how this diversity is generated. One possibility is that relative and net-coding activities are indeed secondarily derived from absolute value representations, which fits the two-stage hierarchical model of decision making. It is noteworthy that past primate studies have demonstrated that OFC neurons encodes values of options in an absolute, or menu-invariant manner (Padoa-Schioppa and Assad 2006; Padoa-Schioppa and Assad 2008) and a recent study in rats showed that the OFC projects densely to the DMS (Schilman, Uylings et al. 2008). This raises the possibility that the OFC may be a source of absolute-value representations to the DMS, although there is still little evidence to support absolute valuation in rodent OFC, and OFC representation may be more tied to goods than to actions (Padoa-Schioppa and Assad 2006). Other major inputs to the DMS include dlPFC and ACC. Previous studies have implicated both dlPFC and ACC in representing state value or tracking average reward rate (Kolling, et al. 2012; reviewed in Lee et al., 2007). It will be

important to see where net value is first calculated and how the representations in cortical and sub-cortical areas differ.

Value input may also be carried by the centromedian/parafascicular nucleus (CM/PF) nucleus in the thalamus. Recordings from CM/PF in a Go/No-Go task showed that the majority of these neurons selectively activated to stimuli predicting small reward, regardless of whether the following action was a Go or No-Go. Further, electrical stimulation of the CM/PF significantly slowed down movements as if the action was towards a low valued, but necessary motion (Minamimoto, Hori et al. 2005). As this nucleus is known to project to the dorsal striatum (Smith, Raju et al. 2004), it may provide sources of signals in the DMS that were negatively modulated by value.

Moreover, dopamine innervation from the SNc or VTA could provide value related information. It would be also worthwhile to examine whether phasic or tonic dopamine levels could be modulated by and/or affect striatal activity.

Lastly, it is also possible that the expression of relative and net-value representations in the DMS does not depend on absolute-value representations, and are *explicitly* or directly represented. It will be crucial to examine whether various value coding neurons interact and how cortical afferents, dopaminergic, and other neurotransmitter systems modulate this diversity of value-modulated responses.

Methods and Additional Analyses

Rats were implanted with custom-made microdrives in the left, anteromedial striatum (1.7 mm anterior to bregma and 2.1 mm lateral to the midline). Extracellular recordings were obtained with twelve independently movable tetrodes using the Cheetah system (Neuralynx) and single units were isolated by manually clustering spike features with MClust (A. D. Redish). Cells were recorded at various depths between 3.5 and 9 mm ventral to the surface of the skull. The boundary between dorsal and ventral striatum was considered to be 6 mm deep (Voorn et al., 2004; Kimchi et al., 2009). The depth of each recorded cell was reconstructed by calculating the number of turns made on each tetrode screw (each turn = 0.32mm) and further confirmed using the final length of each tetrode (through histological examination and measuring the length of the tetrodes after removal of the drive).

Classification of striatal responses

Our analysis focused on the pre-trial initiation period, the 0-300 ms window before odor poke in. The first 10 trials of every block were not used to eliminate the effect of learning.

Next, we regressed each neuron's firing rate by the reward amounts, which varied across blocks.

$$FiringRate = \beta_0 + \beta_L \cdot Q_L + \beta_R \cdot Q_R$$

We used F-test to select for neurons that were statistically significant at the 1% level. The F-test tests whether a proposed regression model as a whole fit the data significantly better than a simpler model (*Firing rate* = β_0). Because this fitting is invariant to the choice of axes, F-test is also invariant to the choice of independent variables.

The amplitude (r) of each neuron was calculated by taking the square root of the sum of the coefficients:

$$r = \sqrt{(\beta_L)^2 + (\beta_R)^2}$$

the polar angle (θ) was calculated by taking the four-quadrant arc-tangent of the coefficients:

$$\theta = \tan^{-1}(\frac{\beta_R}{\beta_L})$$

Similar to the behavioral vector classification, we divided the polarplot into 8 segments of 45^o to classify the neural responses. Neurons whose θ values fall between -22.5^o and +22.5^o were classified as left-positive value-coding, neurons whose θ fall between 22.5^o and 67.5^o were positive-state-value-coding, neurons whose θ fall between 67.5^o and 112.5^o were right-positive value-coding, neurons whose θ fall between 112.5^o and 157.5^o were right-preferring relative-value-coding, etc.

Peri-stimulus time histograms plot the firing rates of each block condition between different time epochs such as odor port entry and water port entry. As the timing between each task event is variable, for visualization purposes, the neural activities were aligned to all

events, and 'time-warped,' so that firing rate (spikes/s) is preserved, but the time window is kept constant (time points are averaged across trials).

More stringent classification of neural responses

Further, we used a more stringent method to classify the responses of our population. By bootstrapping, we obtained a 95% confidence interval on each angle. Here, only neurons whose confidence intervals crossed exactly one classification boundary can be included (each boundary being 0, 45, 90, 135, 180, 225, 270, and 315). With the more stringent classification method, 80/522 neurons were classified as value-coding. Net-value coding neurons was in the majority, comprising 33.8% of the 80 value-coding neurons.

Analysis using the peak window for each neuron

The main analysis window we used was a 300ms time window immediately before odor poke in. Alternatively, we could perform the analysis during each neuron's peak firing. We selected all neurons that had peak firing activity between water-poke-out and odor-poke-in and obtained 181 neurons in total (127/364 in DMS and 54/158 in VS). We performed the above regression analysis and vector projection using a 250ms window around the time of the each neuron's time of peak activity. During this period, the proportion of neurons modulated by block-wise changes in value were not statistically different between DMS and VS (39/127 DMS and 19/54 VS, P=0.55). The dominant neural response in the striatum was correlated with net value (**Fig. 2.12,** left panel). DMS was net-value coding in

both the positive and negative direction (**Fig. 2.12,** middle panel), while in the VS, was primarily negative-net value coding (**Fig. 2.12,** right panel). These results support net value coding over absolute and relative value coding in the striatum during the inter-trial interval.

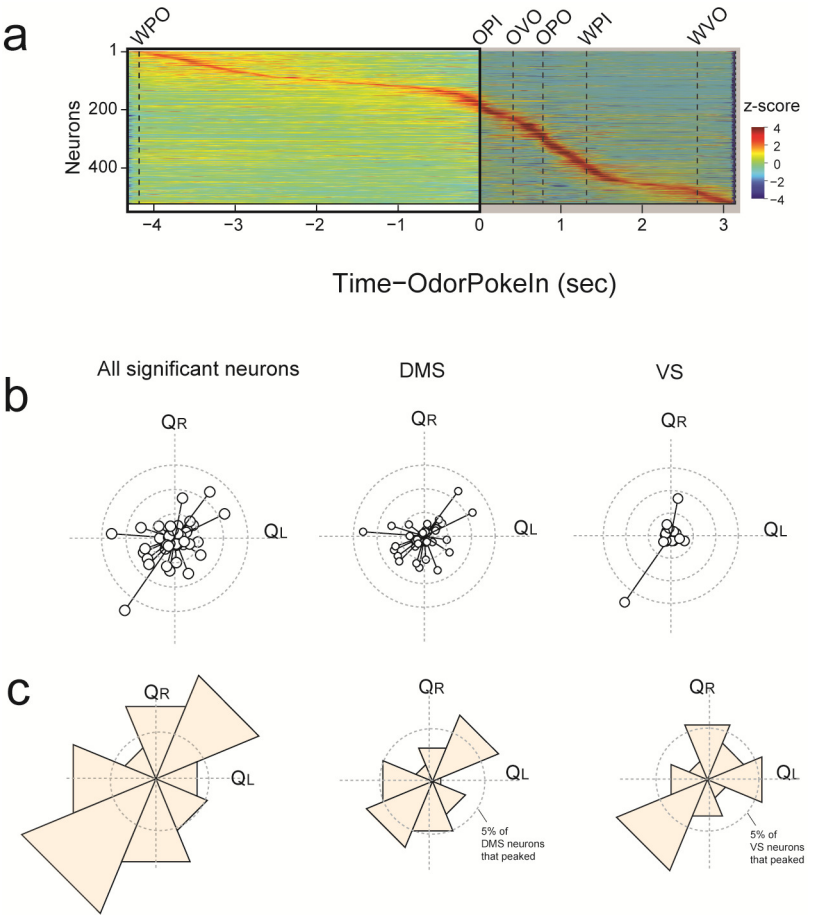

Figure 2.12 Analysis using neurons the peak time window for each neuron
(a) Neurons that peak between WPO and OPI. Same as in Fig. 2.2 but only neurons that show peak firing between WPO and OPI were used (n=181 total). **(b)** Vectors of value coding neurons. Inner ring: r=8, middle ring: r=16, outer ring: r=24. **(c)** Proportion of neurons per value-coding category. Dotted ring: 5% of neurons.

Value-coding cells and their anatomical cell types

Next, we examined how the different types of value-coding neurons may be mapped onto various anatomical cell types. We plotted the firing rate of each neuron against its spike width. (**Fig. 2.13**). Spike width is defined by the interval between the peak and trough of spike waveforms. We show that there are at least two clusters: narrow spike waveform with relatively high firing rates (narrow spikes <200 us, 199 neurons) and wide spike waveforms with relatively low firing rates (wide spikes >200 us, 296 neurons). The latter likely corresponds to medium spiny neurons and the former, tonically active (TAN) neurons and/or fast spiking GABAergic interneurons (Kawaguchi et al., 1993; Gage et al., 2010). We color-coded every value coding neuron on the plot by their coding type (left absolute, right absolute, relative, and net-coding). The proportion of value coding neurons among narrow-spiking and wide-spiking neurons were not statistically different (χ^2 test, $P = 0.13$; 25%, or 50/199 narrow-spiking neurons were value coding and 31% or 93/296 wide-spiking neurons were value coding).

Among the narrow-spiking value coding neurons, there was a non-uniform distribution of value coding types, where net value coding neurons comprised the largest proportion (χ^2 goodness of fit for a uniform distribution, $P = 0.0045$; 8/199, 16/199, 10/199, and 26/199 were left absolute, right absolute, relative, and net value coding, respectively). Yet, among wide-spiking value coding neurons, net value coding types were also the most common, although the distribution of value coding types was not significantly different from uniform (χ^2 , goodness of fit for a uniform distribution, $P = 0.1772$; 22/296,

15/296, 17/296, and 28/296 fell into the respective categories indicated above). Finally, the proportions of absolute, relative, and net value coding neurons between narrow- and wide-spiking neurons were not statistically different from each other (χ^2 test, $P = 0.07$, 0.29, 0.55, and 0.37, for left absolute, right absolute, relative, and state coding neurons, respectively). In sum, we could not find any clear relationship between value coding neural types and putative MSNs or TANs/interneurons. However, these results do not exclude the possibility that different value coding types could be mapped onto D1- or D2-type neurons or to patch/matrix architecture.

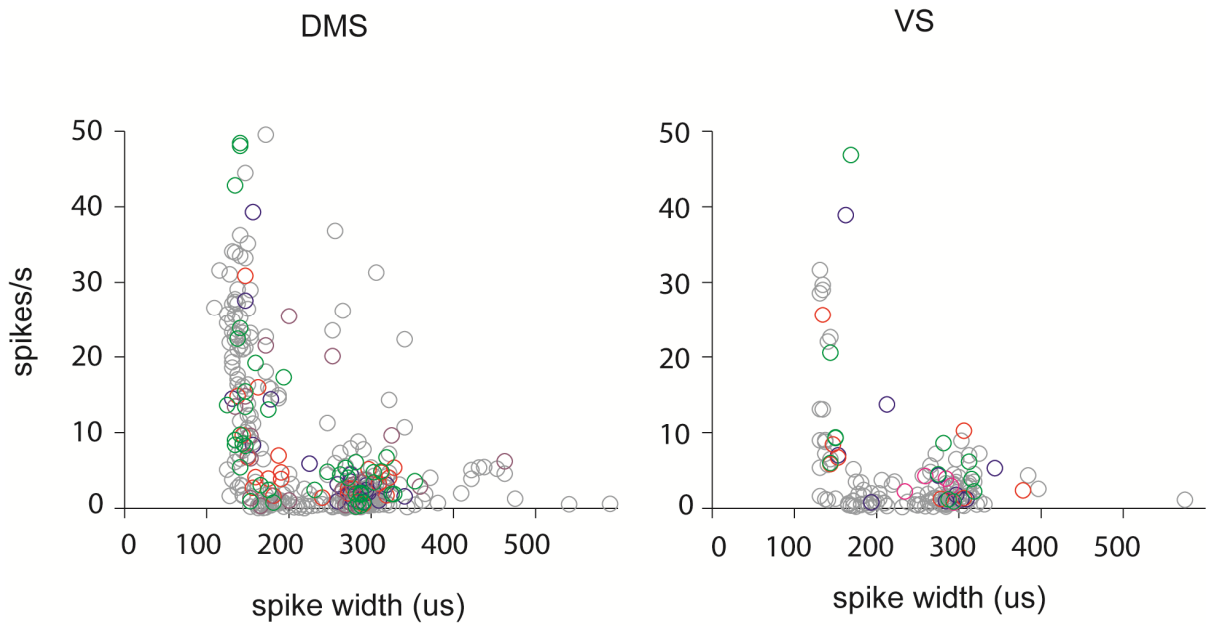

Figure 2.13 Value coding neurons (pre-initiation) and their anatomical cell types
Scatter plot of the spike width and firing rate of every recorded neuron (522 total). Neurons that are color-coded are value-coding and those in grey are non-value coding. Blue: left absolute value; red: right absolute value; purple: relative value; green: net value.

Comparing our proposed vector analysis with previous studies' analyses

In past studies that suggested that DMS primarily encodes absolute action values, firing rates were regressed by the independent variables, QL and QR, representing the 'action value' of left and right, respectively. After running multiple regression, neurons were classified to be absolute-value coding if the coefficients for *either* Q_L or Q_R, but *not both*, were significantly different from zero. Yet, neurons were relative- (or net-) value coding if coefficients for *both* Q_L*and* Q_R were significantly different from zero (**Fig. 2.14a** conventional method type 1). With this regression method, one independent variable is required to be significant for classification of neurons as absolute value-coding, whereas two independent variables are required for classification of neurons as relative or state-value coding.

Conversely, other studies that examined the role of VS and areas such as ACC in relative-value or net-value coding applied the opposite regression using Q_R-Q_L and Q_L+Q_R as independent variables (Ito and Doya, 2009; Seo and Lee, 2007). Different from the above method, only *one* independent variable is required to be significant for neurons to be classified as relative or net-value coding (conventional method type 2).

We first performed a simulation (**Fig. 2.14a**) to test whether the two different methods can correctly capture a neural population that is uniformly involved in all aspects of value coding, i.e. uniformly distributed in the polar coordinate representation (an 'extreme' case). To match numbers obtained in past striatal studies (Samejima et al., 2005; Lau and

Glimcher, 2008), we assumed about 36% of the population was statistically significant at the 5% level. Applying conventional method 1 (regression analysis with Q_L and Q_R), 32% of the population are classified as absolute-value coding, but only 4% are relative- or net-value coding. Therefore, the vast majority (89%) of value coding neurons using this method are classified as absolute-value coding.

Conversely, by applying conventional method 2 (regression analysis with Q_L-Q_R and Q_L+Q_R)(**Fig. 2.14b**), the vast majority, or 89% of significant neurons were relative- or net-value coding, while the rest were absolute-value coding. Therefore, the first conventional method is biased towards classifying neurons into absolute-value coding while the second conventional method is biased towards classification into net or relative value coding types. *Neither* choice of axes for the regression was fair at capturing the simulated uniform distribution. We visualize these systematic biases in classification in **Figure 2.14a-b** (left panels), by plotting a randomly generated and uniformly distributed neural population distributed in two-dimensional space onto each methods' regression plane (neurons falling into different sectors of the critical regions are color-coded and classified accordingly).

By simply replacing the Cartesian coordinate system with a polar coordinate system, so that each neuron's activity is represented by a polar angle (θ) and amplitude (r), equal numbers of neurons are classified into absolute-, relative-, and net-value coding categories (**Fig. 2.14c**). Therefore, analyzing neural responses with our polar method is more fair at classifying changes related to the absolute, relative, and net values of the animal's options.

We obtained similar biases when we applied past regression methods on our own data set (**Fig. 2.15**). We took the 0-300 ms time epoch before trial initiation and and performed the above conventional regression analyses. With conventional method type 1 (regression with Q_L and Q_R), 74.6% of all significant striatal neurons (at the 5% level) were classified as absolute-value coding (**Fig. 2.15**). On the contrary, using conventional method type 2, 79.4% of significant neurons were relative or state coding (**Fig. 2.15**). Lastly, we confirmed the biases by analyzing neural activity under another time window. We examined our data set using another epoch (pre-odor epoch) for the analysis and obtained similar biases with conventional method 1 and 2 (**Fig. 2.15**).

Therefore, discrepancies between our and previous studies may stem from differences in regression methods.

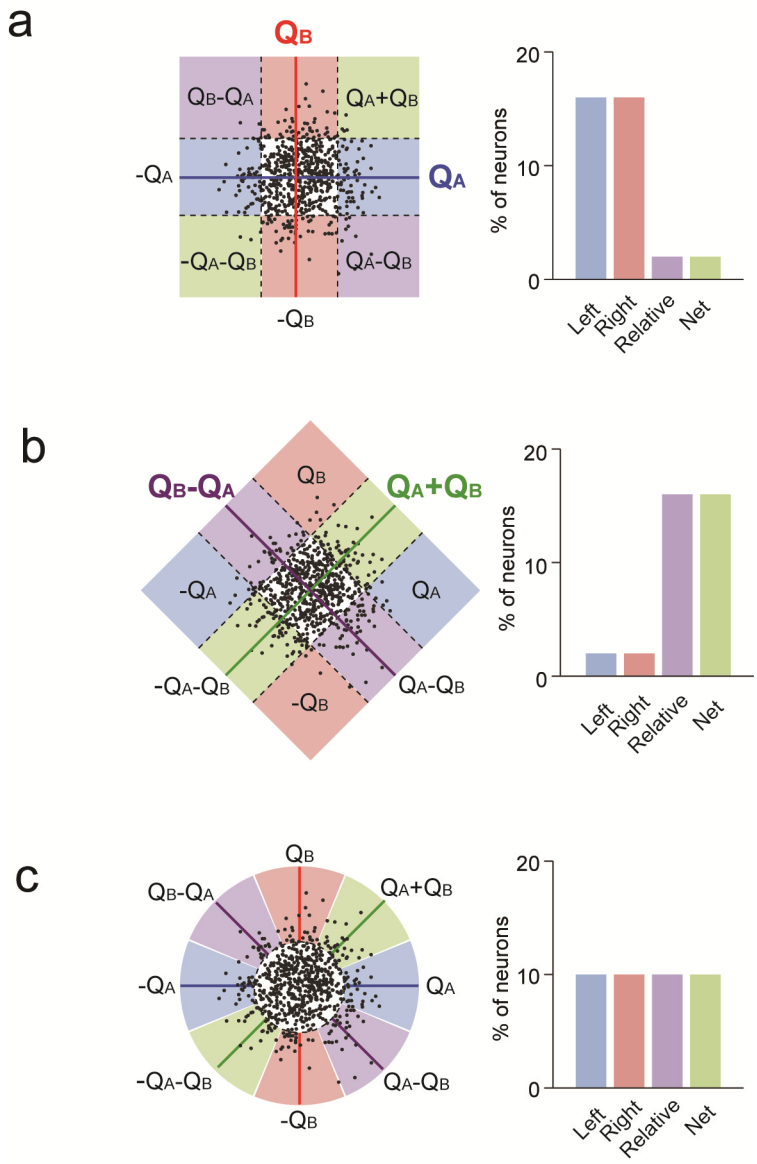

Figure 2.14 Demonstration of classification biases with conventional regression methods using simulation

(a) Conventional method type 1: multiple regression with the independent variables, Q_A (X-axis, blue) and Q_B (Y-axis, red), representing values of A and B, respectively.

(b) Conventional method type 2: multiple regression with the independent variables, Q_B-Q_A (purple) and Q_A+Q_B (green) (tilted axes), representing relative and net values, respectively.

(c) Proposed method: polar representation to avoid bias in classification.

(a-c) Left panels: Black dots represent a randomly generated population of neurons that is uniformlydistributed in two dimensional space (assuming 36% of neurons are significant). These simulated neuronsare projected onto their respective regression planes, where dotted lines represent the critical boundary P=0.05 and critical regions are color-coded.

Right panels: Distribution of simulated absolute, relative, and net value coding types.

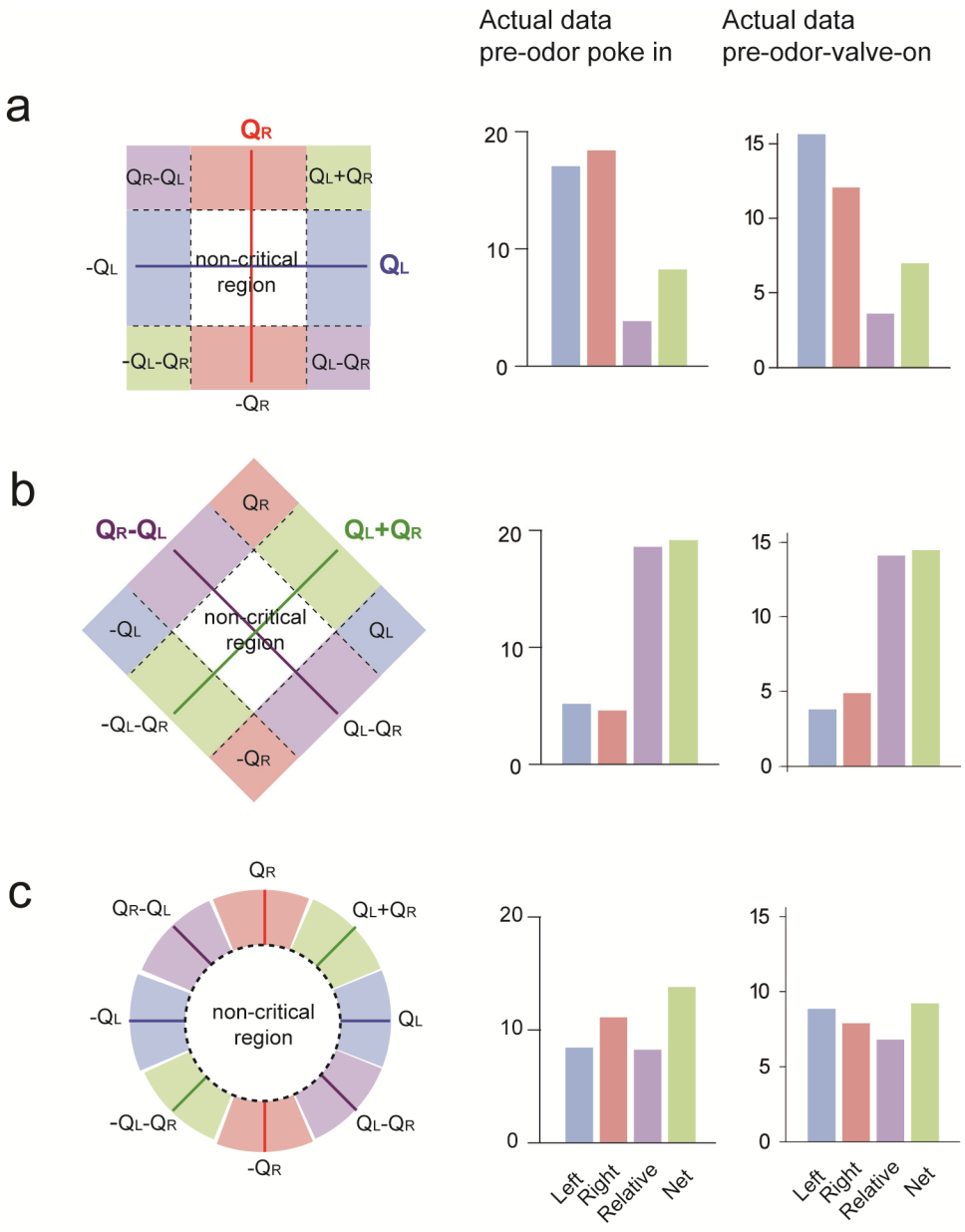

Figure 2.15 Demonstration of classification biases with conventional regression methods using actual data

(a) Conventional method type 1. **(b)** Conventional method type 2. **(c)** Proposed method **(a-c)** Middle panels: Distribution of value coding types on actual data (time epoch: pre-odor-poke-in). Right panels: Distribution of value coding types on actual data (time epoch: pre-odor-valve-on).

Simulations of conventional regression methods: general cases

In the above section, we performed simulations of the biases of classifying absolute, relative, and net value coding neurons of conventional regression methods assuming that the proportion of significantly modulated neurons at the 5% level was about 1/3 of the total population (Samejima et al., Lau and Glimcher, 2009). Below, we demonstrate these biases for more general cases.

Results of categorization by both conventional methods are sensitive to the strengths of neural responses.

In conventional regression methods, the relative proportions of different value coding categories (absolute-value and relative/state coding neurons) are sensitive to the total proportion of statistically significant neurons.

To illustrate this point, we simulate neural populations that are uniformly distributed in two dimensional space with various standard deviations (SD).

When there are no signals in the population, (**Fig. 2.16a**, a normal distribution with SD=1), only chance levels of neurons are selected with a criterion of 5%. In this case, most neurons are significant for a single independent variable (S: single-positive) and very few are significant for *both* independent variables (D: double-positive). On the other hand, in distributions with high SD and thus larger proportions of significant neurons, (**Fig. 2.16**,

SD=6.0), most of these significant neurons are significant for *both* independent variables (D: double-positive) while a smaller fraction are significant for only one independent variable (S: single-positive). Thus, the proportions of neurons classified as single-positive or double positive are sensitive to the SDs of the distributions, or the amount of signal in a population. (Note that with conventional method type 1, single-positive types are absolute-value coding while double-positive types are relative/state coding; the opposite holds for conventional method type 2).

By defining p as the probability of significance for each independent variable, then the probabilities of single-positive, double-positive and total-positive (S+D) types are given as follows (**Fig. 2.16e**):

Single positive: $2p - 2p^2$

Double positive: p^2

Total significant: $2p - p^2$

Therefore, the relative frequency of single-positive points among total positive points is given given (**Fig. 2.16f**):

Single positive/Total: $\frac{2-2p}{2-p}$

Therefore when $p = 0.20$ (20% of the neurons are significantly modulated by either variable), the distribution is about 1.53 SD, so 88.9% of significant neurons are classified as single-positive while only 11.1% are classified as double-positive. (In this case, about 36% of neurons in total are deemed 'significant', which is close to the percentage obtained in the present and previous studies). In Lau and Glimcher (2008) for example, 81% of neurons

were single-positive, 19% were double-positive, and about a third of all neurons were statistically significant.

Contrary to the conventional methods, our method (the polar method) is invariant to overall proportions of statistically significant neurons.

Our proposed method is invariant to the choice of axes for regression

We applied Q_L and Q_R as independent variables for multiple regression. With our proposed method of categorization, 32, 32, 30, and 55 neurons out of 522 in total were left absolute, right absolute, relative, and net value coding, respectively. When we applied Q_L+Q_R and Q_R-Q_L as independent variables, we achieved exactly the same results (to categorize the neurons, all categories are simply shifted +45 degrees). This confirms that our method is invariant to the choice of independent variables (axes).

Figure 2.16: Conventional methods are sensitive to the proportion of significant neurons

(a-d) Simulation of how uniformly distributed response populations at various standard deviations (SD) are classified with conventional methods. (Each black dot represents a neuron). S: Single positive (one independent variable is significant). D: Double positive (two independent variables are significant). Dotted lines represent the critical bounday, P=0.05. **(e)** Curves illustrates how the percent of neurons classified as single-positive (S) and double positive (D) varies as a function of the SD of the distribution. **(f)** Curve illustrates the relative frequency of single-positive (S) neurons.

(e-f) Dotted vertical line represents when the criterion is 1.53 SD (equivalent to having ~36% of the population significant).

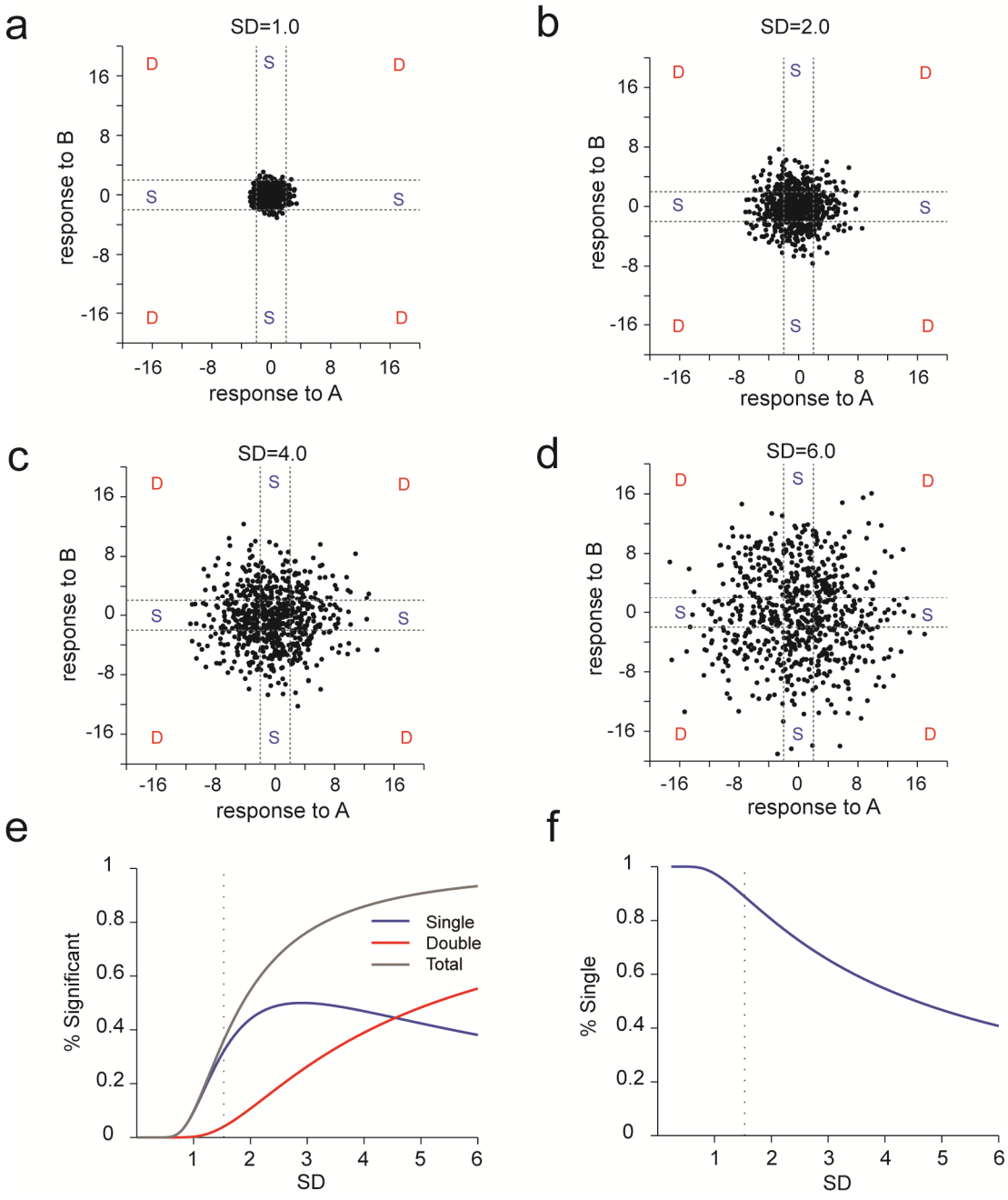

Figure 2.16 (continued).

Conclusion

Two aspects of motivation, the ‘directing’ and ‘energizing’, act in concert to contribute to optimal decision making. In decades past, measurement of response vigor in behavior paradigms such as lever-press tasks and key-pecking tasks had been central, while the study of action selection was not well measured or controlled. In recent years, the opposite problem has emerged, where the primary interest of study is action selection, yet inter-trial intervals are typically controlled by the experimenter, thus ignoring how animals choose whether/how to engage in performing a task. Although there have been a few decision making studies that examined how incentive values of specific goals modulate reaction time, such questions still fall primarily within the 'directing' or outcome-specific effects of motivation (Roesch and Olson 2004; Nakamura and Hikosaka 2006). The present study paves a way towards examining both the directing, or outcome-specific, and the energizing, or outcome-general, effects of motivation on decision making in a unified behavioral task.

The dual-faceted view of motivation stems back to early notions of general and specific drives in motivational theory (Hull 1943; Hebb 1955). Hull imagined that the energizing and directing were two distinct and multiplicative processes that worked together to create motivated behavior. While our results appear to support two dissociable processes, the mechanisms underlying the energizing and directing in our experiments differ substantially from Hull's original conception. Hull imagined that internal, physiological

deprivation activates a singular drive factor that globally energizes actions. Activated by general drive, animals then 'direct' themselves towards specific drive-reducing behaviors, as learned by past stimulus-response reinforcement or habits. Our results in Chapter 1 showed that fluctuations in the net value of external goal options can produce a general drive-like effect without the need to introduce a generalized drive variable. We showed that animal's overall motivation to perform primarily depended on the net value of the animal's current state, and that net value had a long-term or global energizing effect on performance. Tracking the net value of one's state has been hypothesized to be critical for balancing one's overall rewards with the energetic costs of performance and has a key role in models of optimal foraging theory (Pyke 1984; Niv 2007).

Moreover, we provide evidence that the representation of net value may occur in specific brain circuits. We examined the role of the DMS and VS in action selection and motivational drive by selectively lesioning these areas in rats. Lesions of the DMS more profoundly affect the energizing aspect of motivation than the VS. Specifically, rats' initiation time no longer depended on the net value of the block, but rather on the absolute value of the previous trial. The VS-lesioned animals also showed less dependence on net value and more dependence on absolute values after lesions, but to a lesser extent. Moreover, VS lesions caused small but significantly more errors in difficult trials towards relatively better options. Subsequently, we recorded from both DMS and VS and found that during the early epochs of each trial, DMS and VS neurons exhibited changes in firing activity that reflected changes in the animal's values. Our physiology results in the DMS supported our lesion results in that the DMS contained significantly more net value coding neurons than

any other type, as well as more than VS during early trial epochs. The VS contained less value coding neurons in general, although neurons negatively correlated with net value were prominent especially before choice. Moreover, net value representation in the DMS was among the most dominant type of value modulation even throughout the stimulus and choice period.

Net value representation disappeared only at the end of the trial, where DMS neurons primarily encoded the absolute value of the obtained reward, but emerge again before the next trial's initiation. The most prominent activity of VS neurons, on the other hand, occurred after choice. During choice itself, the majority of neurons preferred to encode relative value (for ipsilateral choice). During reward, the VS encoded the absolute value of the obtained reward, similar to the DMS. In all, different from the DMS, value-related activity in the VS is more prominent after choice than it is prior to trial initiation or action selection.

Our results in the DMS extend the current framework of action selection, which states that the DMS primarily encodes the absolute 'action value' of potential choices during early trial epochs. Our results show that the DMS is not restricted to encoding absolute action values and that it predominantly encodes net value. Although previous studies used different animal models (primates instead of rodents), task structure (varying reward value by temporal discounting or reward probability vs. varying reward size), and analysis epochs (after stimulus onset instead of before trial initiation), we show that the critical differences in our results may stem primarily from our analysis methods (Chapter 2, methods). Our polar

vector quantification of value modulation allows comprehensive and unbiased classification of neural activity related to absolute, relative, and net coding. In sum, our results suggest that the DMS, rather than the VS, is more closely linked to net value coding and the flexible regulation of motivational drive than previously thought.

Furthermore, the striatum is heavily interconnected with cortex, other basal ganglia nuclei, dopamine neurons, and other brain regions, many of which exhibit value-related signals. Future studies will need to address how signals from these various interconnected brain areas differ from one another or differ in their effects on projection targets. For example, how do the tonic or phasic responses of dopamine neurons contribute to the maintenance or representation of net value throughout the course of states of high net value? How does the DMS quickly adapt to changes in net value and how do DMS' signals affect downstream motor resources to allow switching between exploiting the task and exploring other behaviors? Lastly, we recognize that the regulation of motivational drive itself is multifaceted and probably cannot be fully captured by computing the net value of options alone. Future studies will need to address more closely the interactions between net value representation and areas signaling energetic costs, physiological deprivation or satiation, or other factors such as circadian-dependent arousal levels.

Choosing the general pace of performance (or whether to perform at all) in conjunction with what specific action to take is vital for behavioral regulation. By providing a theoretical framework, a behavioral paradigm, and analytical tools, our study instigates a more inclusive understanding of decision making. Applying these approaches to future

experiments in different brain regions will further our understanding of how the brain regulates value representation and goal-oriented behavior.

References

Aberman, J. E. and J. D. Salamone (1999). "Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement." Neuroscience **92**(2): 545-552.

Alexander, G. E. and M. D. Crutcher (1990). "Functional architecture of basal ganglia circuits: neural substrates of parallel processing." Trends in Neurosciences **13**(7): 266-271.

Bailey, K. R. and R. G. Mair (2006). "The Role of Striatum in Initiation and Execution of Learned Action Sequences in Rats." The Journal of Neuroscience **26**(3): 1016-1025.

Balleine, B. W. and A. Dickinson (1998). "Goal-directed instrumental action: contingency and incentive learning and their cortical substrates." Neuropharmacology **37**(4–5): 407-419.

Barto, A. G., R. S. Sutton, et al. (1983). "Neuronlike adaptive elements that can solve difficult learning control problems." IEEE Transactions on Systems, Man, & Cybernetics **13**(5): 834-846.

Berridge, K. C. and T. E. Robinson (1998). "What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience?" Brain Research Reviews **28**(3): 309-369.

Bolles, R. C. (1975). Theory of motivation. New York, Harper & Row.

Bornstein, A. M. and N. D. Daw (2011). "Multiplicity of control in the basal ganglia: computational roles of striatal subregions." Current Opinion in Neurobiology **21**(3): 374-380.

Brown, V. and T. Robbins (1989). "Elementary processes of response selection mediated by distinct regions of the striatum." The Journal of Neuroscience **9**(11): 3760-3765.

Cai, X., S. Kim, et al. (2011). "Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice." Neuron **69**(1): 170-182.

Cardinal, R. N., J. A. Parkinson, et al. (2002). "Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats." Behavioral Neuroscience **116**(4): 553-567.

Cardinal, R. N., D. R. Pennicott, et al. (2001). "Impulsive Choice Induced in Rats by Lesions of the Nucleus Accumbens Core." Science **292**(5526): 2499-2501.

Cousins, M. S., A. Atherton, et al. (1996). "Nucleus accumbens dopamine depletions alter relative response allocation in a T-maze cost/benefit task." Behavioural Brain Research **74**(1–2): 189-197.

Croxson, P. L., M. E. Walton, et al. (2009). "Effort-Based Cost–Benefit Valuation and the Human Brain." The Journal of Neuroscience **29**(14): 4531-4541.

Cury, K. M. and N. Uchida (2010). "Robust Odor Coding via Inhalation-Coupled Transient Activity in the Mammalian Olfactory Bulb." Neuron **68**(3): 570-585.

Daw, N. D. and K. Doya (2006). "The computational neurobiology of learning and reward." Current Opinion in Neurobiology **16**(2): 199-204.

Delgado, M. R., V. A. Stenger, et al. (2004). "Motivation-dependent Responses in the Human Caudate Nucleus." Cerebral Cortex **14**(9): 1022-1030.

Desban, M., M. L. Kemel, et al. (1993). "Spatial organization of patch and matrix compartments in the rat striatum." Neuroscience **57**(3): 661-671.
Dickinson, A. and B. Balleine (1994). "Motivational control of goal-directed action." Learning & Behavior **22**(1): 1-18.
Dorris, M. C. and P. W. Glimcher (2004). "Activity in Posterior Parietal Cortex Is Correlated with the Relative Subjective Desirability of Action." Neuron **44**(2): 365-378.
Everitt, B. J. and T. W. Robbins (2005). "Neural systems of reinforcement for drug addiction: from actions to habits to compulsion." Nat Neurosci **8**(11): 1481-1489.
Ferster, C. B. and B. F. Skinner (1957). Schedules of reinforcement. East Norwalk, CT, US, Appleton-Century-Crofts.
Gerfen, C. R. (1992). "The Neostriatal Mosaic: Multiple Levels of Compartmental Organization in the Basal Ganglia." Annual Review of Neuroscience **15**(1): 285-320.
Graybiel, A. M. and C. W. Ragsdale (1978). "Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining." Proceedings of the National Academy of Sciences **75**(11): 5723-5726.
Gurney, K., T. J. Prescott, et al. (2001). "A computational model of action selection in the basal ganglia. I. A new functional anatomy." Biological Cybernetics **84**(6): 401-410.
Haber, S. N., J. L. Fudge, et al. (2000). "Striatonigrostriatal Pathways in Primates Form an Ascending Spiral from the Shell to the Dorsolateral Striatum." J. Neurosci. **20**(6): 2369-2382.
Hall, J., J. A. Parkinson, et al. (2001). "Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour." European Journal of Neuroscience **13**(10): 1984-1992.
Hauber, W. and W. J. Schmidt (1994). "Differential effects of lesions of the dorsomedial and dorsolateral caudate-putamen on reaction time performance in rats." Behavioural Brain Research **60**(2): 211-215.
Hayden, B. Y., J. M. Pearson, et al. (2011). "Neuronal basis of sequential foraging decisions in a patchy environment." Nat Neurosci **14**(7): 933-939.
Hebb, D. O. (1955). "Drives and the C. N. S. (conceptual nervous system)." Psychological Review **62**(4): 243-254.
Herrnstein, R. J. (1970). "On the law of effect." Journal of the Experimental Analysis of Behavior **13**(2): 243-266.
Houk, J. C., J. L. Adams, et al. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. Models of information processing in the basal ganglia. J. C. Houk, J. L. Davis and D. G. Beiser. Cambridge, MA, US, The MIT Press**:** 249-270.
Hull, C. L. (1943). Principles of behavior. New York, Appleton-Century-Crofts.
Ikemoto, S. and J. Panksepp (1999). "The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking." Brain Research Reviews **31**(1): 6-41.
Ito, M. and K. Doya (2009). "Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia." J. Neurosci. **29**(31): 9861-9874.
Ito, R., J. W. Dalley, et al. (2002). "Dopamine Release in the Dorsal Striatum during Cocaine-Seeking Behavior under the Control of a Drug-Associated Cue." The Journal of Neuroscience **22**(14): 6247-6253.

Joel, D. and I. Weiner (2000). "The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum." Neuroscience **96**(3): 451-474.
Joyce, J. N., D. W. Sapp, et al. (1986). "Human striatal dopamine receptors are organized in compartments." Proceedings of the National Academy of Sciences **83**(20): 8002-8006.
Kable, J. W. and P. W. Glimcher (2009). "The Neurobiology of Decision: Consensus and Controversy." Neuron **63**(6): 733-745.
Kawagoe, R., Y. Takikawa, et al. (1998). "Expectation of reward modulates cognitive signals in the basal ganglia." Nat Neurosci **1**(5): 411-416.
Kelley, A. E. (2004). "Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning." Neuroscience & Biobehavioral Reviews **27**(8): 765-776.
Killcross, S. and E. Coutureau (2003). "Coordination of Actions and Habits in the Medial Prefrontal Cortex of Rats." Cerebral Cortex **13**(4): 400-408.
Killeen, P. R., S. J. Hanson, et al. (1978). "Arousal: Its genesis and manifestation as response rate." Psychological Review **85**(6): 571-581.
Kim, H., J. H. Sul, et al. (2009). "Role of Striatum in Updating Values of Chosen Actions." J. Neurosci. **29**(47): 14701-14712.
Kim, S., J. Hwang, et al. (2008). "Prefrontal Coding of Temporally Discounted Values during Intertemporal Choice." Neuron **59**(1): 161-172.
Kimchi, E. Y. and M. Laubach (2009). "Dynamic Encoding of Action Selection by the Medial Striatum." The Journal of Neuroscience **29**(10): 3148-3159.
Kolling, N., T. E. J. Behrens, et al. (2012). "Neural Mechanisms of Foraging." Science **336**(6077): 95-98.
Kravitz, A. V., B. S. Freeze, et al. (2010). "Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry." Nature **466**(7306): 622-626.
Lau, B. and P. W. Glimcher (2008). "Value Representations in the Primate Striatum during Matching Behavior." Neuron **58**(3): 451-463.
Lauwereyns, J., K. Watanabe, et al. (2002). "A neural correlate of response bias in monkey caudate nucleus." Nature **418**(6896): 413-417.
Le Moine, C. and B. Bloch (1995). "D1 and D2 dopamine receptor gene expression in the rat striatum: Sensitive cRNA probes demonstrate prominent segregation of D1 and D2 mRNAS in distinct neuronal populations of the dorsal and ventral striatum." The Journal of Comparative Neurology **355**(3): 418-426.
Lee, D., H. Seo, et al. (2012). "Neural Basis of Reinforcement Learning and Decision Making." Annual Review of Neuroscience **35**(1): null.
McClure, S. M., N. D. Daw, et al. (2003). "A computational substrate for incentive salience." Trends in Neurosciences **26**(8): 423-428.
Middleton, F. A. and P. L. Strick (2000). "Basal ganglia and cerebellar loops: motor and cognitive circuits." Brain Research Reviews **31**(2–3): 236-250.
Minamimoto, T., Y. Hori, et al. (2005). "Complementary Process to Response Bias in the Centromedian Nucleus of the Thalamus." Science **308**(5729): 1798-1801.
Mogenson, G. J., D. L. Jones, et al. (1980). "From motivation to action: Functional interface between the limbic system and the motor system." Progress in Neurobiology **14**(2-3): 69-97.

Mogenson, G. J., M. Takigawa, et al. (1979). "Self-stimulation of the nucleus accumbens and ventral tegmental area of tsai attenuated by microinjections of spiroperidol into the nucleus accumbens." Brain Research **171**(2): 247-259.
Nakamura, K. and O. Hikosaka (2006). "Role of Dopamine in the Primate Caudate Nucleus in Reward Modulation of Saccades." J. Neurosci. **26**(20): 5360-5369.
Nakano, K., T. Kayahara, et al. (2000). "Neural circuits and functional organization of the striatum." Journal of Neurology **247**(0): V1-V15.
Nicholaidis, S. and N. Rowland (1974). "Long-term self-intravenous "drinking" in the rat." Journal of Comparative and Physiological Psychology **87**(1): 1-15.
Niv, Y. (2007). "Cost, Benefit, Tonic, Phasic." Annals of the New York Academy of Sciences **1104**(1): 357-376.
Niv, Y., N. Daw, et al. (2007). "Tonic dopamine: opportunity costs and the control of response vigor." Psychopharmacology **191**(3): 507-520.
Niv, Y., D. Joel, et al. (2006). "A normative perspective on motivation." Trends in Cognitive Sciences **10**(8): 375-381.
O'Doherty, J., P. Dayan, et al. (2004). "Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning." Science **304**(5669): 452-454.
Olds, J. and P. Milner (1954). "POSITIVE REINFORCEMENT PRODUCED BY ELECTRICAL STIMULATION OF SEPTAL AREA AND OTHER REGIONS OF RAT BRAIN." Journal of Comparative and Physiological Psychology **47**(6): 419-427.
Opris, I., M. Lebedev, et al. (2011). "Motor Planning under Unpredictable Reward: Modulations of Movement Vigor and Primate Striatum Activity." Frontiers in Neuroscience **5**.
Padoa-Schioppa, C. (2011). "Neurobiology of Economic Choice: A Good-Based Model." Annual Review of Neuroscience **34**(1): null.
Padoa-Schioppa, C. and J. A. Assad (2006). "Neurons in the orbitofrontal cortex encode economic value." Nature **441**(7090): 223-226.
Padoa-Schioppa, C. and J. A. Assad (2008). "The representation of economic value in the orbitofrontal cortex is invariant for changes of menu." Nat Neurosci **11**(1): 95-102.
Palmiter, R. D. (2008). "Dopamine Signaling in the Dorsal Striatum Is Essential for Motivated Behaviors." Annals of the New York Academy of Sciences **1129**(1): 35-46.
Parent, A. and L.-N. Hazrati (1995). "Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop." Brain Research Reviews **20**(1): 91-127.
Platt, M. L. (2002). "Neural correlates of decisions." Current Opinion in Neurobiology **12**(2): 141-148.
Platt, M. L. and P. W. Glimcher (1999). "Neural correlates of decision variables in parietal cortex." Nature **400**(6741): 233-238.
Pyke, G. H. (1984). "Optimal Foraging Theory: A Critical Review." Annual Review of Ecology and Systematics **15**(1): 523-575.
Ragozzino, M. E. (2007). "The Contribution of the Medial Prefrontal Cortex, Orbitofrontal Cortex, and Dorsomedial Striatum to Behavioral Flexibility." Annals of the New York Academy of Sciences **1121**(1): 355-375.
Rangel, A., C. Camerer, et al. (2008). "A framework for studying the neurobiology of value-based decision making." Nat Rev Neurosci **9**(7): 545-556.

Rescorla, R. A. (1971). "Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning." Learning and Motivation **2**(2): 113-123.
Reuter, J., T. Raedler, et al. (2005). "Pathological gambling is linked to reduced activation of the mesolimbic reward system." Nat Neurosci **8**(2): 147-148.
Reynolds, J. N. J. and J. R. Wickens (2002). "Dopamine-dependent plasticity of corticostriatal synapses." Neural Networks **15**(4–6): 507-521.
Robinson, S., A. Rainwater, et al. (2007). "Viral restoration of dopamine signaling to the dorsal striatum restores instrumental conditioning to dopamine-deficient mice." Psychopharmacology **191**(3): 567-578.
Roesch, M. R. and C. R. Olson (2004). "Neuronal Activity Related to Reward Value and Motivation in Primate Frontal Cortex." Science **304**(5668): 307-310.
Rorie, A. E., J. Gao, et al. (2010). "Integration of Sensory and Reward Information during Perceptual Decision-Making in Lateral Intraparietal Cortex (LIP) of the Macaque Monkey." PLoS One **5**(2): e9308.
Salamone, J., M. Correa, et al. (2007). "Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits." Psychopharmacology **191**(3): 461-482.
Salamone, J. D. and M. Correa (2002). "Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine." Behavioural Brain Research **137**(1–2): 3-25.
Salamone, J. D., A. Wisniecki, et al. (2001). "Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement." Neuroscience **105**(4): 863-870.
Samejima, K., Y. Ueda, et al. (2005). "Representation of Action-Specific Reward Values in the Striatum." Science **310**(5752): 1337-1340.
Scheres, A., M. P. Milham, et al. (2007). "Ventral Striatal Hyporesponsiveness During Reward Anticipation in Attention-Deficit/Hyperactivity Disorder." Biological Psychiatry **61**(5): 720-724.
Schilman, E. A., H. B. M. Uylings, et al. (2008). "The orbital cortex in rats topographically projects to central parts of the caudate-putamen complex." Neuroscience Letters **432**(1): 40-45.
Schultz, W. (1998). "Predictive Reward Signal of Dopamine Neurons." Journal of Neurophysiology **80**(1): 1-27.
Selemon, L. and P. Goldman-Rakic (1985). "Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey." The Journal of Neuroscience **5**(3): 776-794.
Seo, H., D. J. Barraclough, et al. (2009). "Lateral Intraparietal Cortex and Reinforcement Learning during a Mixed-Strategy Game." The Journal of Neuroscience **29**(22): 7278-7289.
Seo, H. and D. Lee (2007). "Temporal Filtering of Reward Signals in the Dorsal Anterior Cingulate Cortex during a Mixed-Strategy Game." J. Neurosci. **27**(31): 8366-8377.
Seo, M., E. Lee, et al. (2012). "Action Selection and Action Value in Frontal-Striatal Circuits." Neuron **74**(5): 947-960.
Smith, Y., D. V. Raju, et al. (2004). "The thalamostriatal system: a highly specific network of the basal ganglia circuitry." Trends in Neurosciences **27**(9): 520-527.

Sotak, B. N., T. S. Hnasko, et al. (2005). "Dysregulation of dopamine signaling in the dorsal striatum inhibits feeding." Brain Research **1061**(2): 88-96.
Sugrue, L. P., G. S. Corrado, et al. (2005). "Choosing the greater of two goods: neural currencies for valuation and decision making." Nat Rev Neurosci **6**(5): 363-375.
Sul, J. H., H. Kim, et al. (2010). "Distinct Roles of Rodent Orbitofrontal and Medial Prefrontal Cortex in Decision Making." Neuron **66**(3): 449-460.
Sutton, R. S. and A. G. Barto (1998). Introduction to Reinforcement Learning, MIT Press.
Szczypka, M. S., K. Kwok, et al. (2001). "Dopamine production in the caudate putamen restores feeding in dopamine-deficient mice." Neuron **30**(3): 819-828.
Thorn, C. A., H. Atallah, et al. (2010). "Differential Dynamics of Activity Changes in Dorsolateral and Dorsomedial Striatal Loops during Learning." Neuron **66**(5): 781-795.
Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases." Science **185**(4157): 1124-1131.
Uchida, N. and Z. F. Mainen (2003). "Speed and accuracy of olfactory discrimination in the rat." Nat Neurosci **6**(11): 1224-1229.
Vanderschuren, L. J. M. J., P. Di Ciano, et al. (2005). "Involvement of the Dorsal Striatum in Cue-Controlled Cocaine Seeking." The Journal of Neuroscience **25**(38): 8665-8670.
Volkow, N. D., G.-J. Wang, et al. (2002). "“Nonhedonic” food motivation in humans involves dopamine in the dorsal striatum and methylphenidate amplifies this effect." Synapse **44**(3): 175-180.
Volkow, N. D., G.-J. Wang, et al. "Cocaine Cues and Dopamine in Dorsal Striatum: Mechanism of Craving in Cocaine Addiction." The Journal of Neuroscience **26**(24): 6583-6588.
Von Neumann, J. and O. Morgenstern (1944). Theory of games and economic behavior. Princeton, NJ, US, Princeton University Press.
Walker, R. H., G. W. Arbuthnott, et al. (1993). "Dendritic domains of medium spiny neurons in the primate striatum: Relationships to striosomal borders." The Journal of Comparative Neurology **337**(4): 614-628.
Walton, M. E., D. M. Bannerman, et al. (2003). "Functional Specialization within Medial Frontal Cortex of the Anterior Cingulate for Evaluating Effort-Related Decisions." The Journal of Neuroscience **23**(16): 6475-6479.
Walton, M. E., S. W. Kennerley, et al. (2006). "Weighing up the benefits of work: Behavioral and neural analyses of effort-related decision making." Neural Networks **19**(8): 1302-1314.
Wong, D. F., H. Kuwabara, et al. (2006). "Increased Occupancy of Dopamine Receptors in Human Striatum during Cue-Elicited Cocaine Craving." Neuropsychopharmacology **31**(12): 2716-2727.
Wunderlich, K., A. Rangel, et al. (2009). "Neural computations underlying action-based decision making in the human brain." Proceedings of the National Academy of Sciences **106**(40): 17199-17204.
Wunderlich, K., A. Rangel, et al. (2010). "Economic choices can be made using only stimulus values." Proceedings of the National Academy of Sciences **107**(34): 15005-15010.

Yin, H. H., B. J. Knowlton, et al. (2004). "Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning." European Journal of Neuroscience **19**(1): 181-189.
Yin, H. H., B. J. Knowlton, et al. (2005). "Blockade of NMDA receptors in the dorsomedial striatum prevents action–outcome learning in instrumental conditioning." European Journal of Neuroscience **22**(2): 505-512.
Yin, H. H., B. J. Knowlton, et al. (2006). "Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning." Behavioural Brain Research **166**(2): 189-196.
Yin, H. H., S. B. Ostlund, et al. (2005). "The role of the dorsomedial striatum in instrumental conditioning." European Journal of Neuroscience **22**(2): 513-523.